一种可以实现真正通用人工智能的新方案和具体实施步骤

陈永聪[1*] 曾婷[2] 张俊[3]
1，新千年未来科技（北京）有限公司，北京市，100084, E-mail: yongcongchen@sina.com；
2，清华大学图书馆，北京市，100084， E-mail: ceng-t@mail.tsinghua.edu.cn；
3，沈阳师范大学，沈阳市，110034

摘 要：

目前主流的人工智能，普遍采用“注意力机制 + 深度学习”+“强化学习”的技术道路。在 AIGC（Artificial Intelligence Generated Content）领域取得了长足进步，掀起了大模型的技术浪潮[2][13]。但在那些需要和实际环境互动的领域，比如老人护理，家庭保姆，农业生产，车辆驾驶等领域，试错成本很高，需要大量试错的强化学习过程难以实现。所以，要想实现能适用于任何领域的通用人工智能，我们既要利用现有技术，又要解决现有技术的缺陷，从而推动人工智能的技术浪潮进一步发展。在本文中，我们分析了大模型技术路线的局限性，并针对这些局限性，提出了解决方案，从而解决了大模型的固有缺陷。在本文中，我们将揭示如何一步一步实现通用人工智能。

关键词：通用人工智能 AGI 强化学习 大模型 ChatGPT GPT-4

1，引言

目前主流的人工智能大模型，带来了通用人工智能的火花[1]，但它还不是真正意义上的通用人工智能。目前人工智能大模型的能力上限在哪里？“注意力机制 + 深度学习”+“强化学习”能实现真正的“通用人工智能”吗？我们认为目前人工智能大模型无法解决下面的严重缺陷：

1.1，不能自主解决问题。

比如目前人工智能，它看到主人摔倒时，并不会主动过来帮忙[7]。这是因为机器没有自己的需求，就不可能产生自己的目标。由于机器没有自己的目标，就不可能主动创建一个任务。也就是说，大模型不会自主创建新的程序流程！

大模型本质就是一种编程平台。使用的编程语言就是自然语言[14]。所以，无论我们添加多少高级函数到大模型中去，也无论我们集成多少工具、APP 到大模型中去，大模型都不会自发地去创建新的流程。它所有的流程都是预设的，要么来自于程序预设，要么来自于数据统计。这两种方式，本质上都是“使用预置的流程来处理所有问题”[8][9][10][11][12]。无论这个流程中有多少 if…else…，考虑到多少种可能性，它都是预置的，预先就存在的。它不是针对具体任务，机器自我创造出来的！所以，按照预定流程决策的机器，就是“书呆子”型机器智能，决策无法灵活变通，难以面对实际的社会生活中层出不穷的意外情况，这也是目前人工智能的窘境。

1.2，知识无法实时更新。

目前人工智能，采用大数据训练，知识无法实现实时更新。而知识的实时更新，对于和环境互动的机器而言，至关重要。因为机器和环境的互动，就是机器

获得新知识的过程。如果机器所获得的知识无法实时更新，就会导致机器无法实时根据环境的反馈来更新自己的决策知识。所以，这样的机器，面对相同的输入信息，就会不断犯相同的错误[3]。

1.3，无法适用于需要和真实环境互动的领域。

在需要和真实环境互动的领域，比如自动驾驶、做家务、护理病人等领域，机器需要建立自己行为和外界环境之间的互动决策知识。而这些领域难以大量地试错，所以机器无法通过强化学习，在真实的环境中通过互动来建立这些领域的决策知识[2][4]。

我们希望，未来每个家庭，都有一个机器保姆；所有车辆，都能自动驾驶；机器人承担所有的工业、农业、服务业，人类主要工作就是享受生活的美好。但目前人工智能的技术方案，还无法实现上述场景。

2，如何创建知识？

2.1，如何描述一个矩阵包含的信息？

虽然一个矩阵可能包含很多信息，但我们可以通过建立一套坐标基底簇，来表达矩阵中所有信息。如果这套坐标基底簇是完备的，是正交的，就可以最简洁地描述矩阵中所有信息。如果我们建立的坐标基底簇并非正交的，但是完备的，那么我们同样可以用这套坐标基底簇来表达矩阵中任意信息。如果坐标基底簇是非完备的，那么矩阵中就存在一些矢量，无法通过这套坐标基底簇来表达，这时我们就需要增加坐标基底簇的维度。

如果基底坐标簇是本征正交基底，那么我们实现了用最简洁的系数来表达这个矢量的全部信息。如果基底坐标簇不是完全正交的，那么我们希望它尽可能接近正交基底簇，因为这时我们获得的系数矩阵是稀疏的（高效表达）。但如果我们只关心矩阵中部分常见信息，我们就可以用常见信息模式作为坐标基底，这样的基底，对整体信息而言，不是高效的表达方式，但对于那些常用信息而言，则是一种高效的表达方式（系数矩阵是稀疏的）。所以，如果我们生活在一个信息矩阵空间中，当我们需要识别、分析和生成各种各样的信息时，最重要的就是：找到信息矩阵空间中一套基底坐标簇。

2.2 人类是如何创建知识的？

人类能识别的信息只是我们世界中信息的极小一部分。这是因为，我们人类对信息的分辨率是有限的。一颗小草上 A 原子和 B 原子排列的相对时空关系，也是一种信息，但我们不会去识别它。

所以人类在进化的过程中，产生了 Tokens 识别能力。Tokens 就是人类常用的最小信息单元，比如一根直线。Tokens 本身就是一种“世界模型”，它是人类用于搭建宏伟的知识殿堂的最小“世界模型”。人类在进化过程中，形成了采用 Tokens 这样的“模型”来识别周围信息的“模式识别”能力，极大的提升了信息识别的能效比。这是进化带给我们的礼物。

所以我们把人类习惯使用的最小信息单元，比如点、线、面、颜色、纹理、曲度、音节、音调、符号、触觉、温度、方向等作为 Tokens，那么我们人类就生活在一个由 Tokens 组成的 4 维度矩阵中（三维空间+时间维度）。对人类而言，从宇宙大爆炸到今天，这个 4 维 Tokens 矩阵，就包含了全部知识。

人类对其中的常见 Tokens 组合，慢慢使用某种符号（语言符号）来代表，这就是概念。人类使用概念，来描述矩阵中的任意信息（矢量）：就是聊天、写文

章。而这些概念，就是我们所处信息空间矩阵中的一套坐标基底簇。

在这样的基底簇下，常见信息（矢量）的系数表达就是稀疏的。比如“投资人”代表“人属科，有钱，想赚更多的钱，找人帮他赚，承担风险，签协议，分享收益...”。

概念包含常见 Tokens 组合，也包含语言符号。而且由于语言符号更加频繁的出现，它代表性更高，可能成为一个概念最常用的入口。

显然，人类的概念，并非正交系。人类习惯于把经常出现的 Tokens 组合，作为一个概念。概念所包含的 Tokens 组合，可能存在不重叠、部分重叠和完全包含等关系。那些存在于大量事物中的共有 Tokens，代表性高，但分辨率低，数量更少，它们代表抽象概念。在抽象概念的基础上，增加更多 Tokens，形成更加具体的概念，其代表的范围缩小，分辨率更高。

虽然这样的坐标基底簇，表达全部信息时，效率并不高。但它们可以高效地表达常见信息。比如“猫”、“狗”这样的概念，可能存在大量的共有 Tokens，它们是非正交的。但对人类表达日常的信息时，却是高效的。

而且，这对信息的泛化至关重要。因为事物的属性，本质上是有组成它的 Tokens 属性组合而成。比如“猫”是一种常见的 Tokens 在空间和时间上的排列方式，这个排列方式中可能包含“猫”的语言、文字、声音、图像、动作、触觉等等多模态矩阵信息元素。这个排列方式中，部分矩阵元素可能拥有更高的权重，因为它们更加常见，它们可能都属于“动物”这个概念。“动物”包含的元素更少，其适用范围就更大，所以在“猫”和“狗”之间，通过它们共有的 Tokens（比如和“动物”这个概念相关的 Tokens）连接的属性就可以直接复用。这就是信息泛化过程，也是智能的起源。

2.3 深度学习的工作原理是什么？

深度学习中，每一层神经网络的系数，背后是一组隐含的坐标基底。A 层神经网络到 A+1 层神经网络，本质是 A 层系数矩阵（和 A 层对应的隐含坐标基底簇一起表达了信息）到 B 层系数矩阵（和 B 层对应的隐含坐标基底簇一起表达了信息）的一次基底变换过程。然后通过非线性函数，对信息作部分压缩或者抛弃。深度学习的本质是使用“试错法”，寻找到一套合适的坐标基底簇，可以让输入信息中的“有用信息”系数矩阵稀疏化。

残差网络的目的是减小每一层神经网络的信息损失量，使得机器可以进行多次变换，从而有更大概率找到优选基底簇。正则化的目的，是让中间层神经网络的隐含基底尽可能靠近正交系，这样避免出现维度之间的彼此影响，从而避免出现局部最优点。它通过迫使中间层的系数矩阵靠近稀疏化矩阵来实现这一目的。

深度学习所创建的高维特征，也就是它的信息矩阵中的一套坐标基底簇。但它没有“使用常见 Tokens 组合”这个约束条件。它使用误差约束下的“试错法”，建立的坐标基底簇，更加倾向于高效表达的正交系。它对整体有用信息表达效率更高，但和人类习惯不一样（人类只需要高效表达那些常见信息），所以“深度学习”和“人类”交流就是“鸡同鸭讲”，双方说不到一块。

而大模型中，注意力机制，本质就是通过预训练获得的局部统计知识，来预测特定 Tokens 组合的常见程度（出现概率），并以它们成为优选的坐标基底簇，来识别、分析和生成各种 Tokens 组合。

这样的基底簇，更加符合人类的习惯。所以大模型和人类之间可以实现语言交流。这样的整体表达效率不一定高，但对常见信息的表达效率更高。

2.4 注意力机制的本质是什么？

注意力机制的核心，本质是一种贝叶斯推理（条件概率）。可以概况为“已知 N 个 Tokens 组合出现的概率，求 M 个 Tokens 组合出现的概率”。

在人类的语言中，N 个 Tokens 的组合几乎无穷无尽，而 M 个 Tokens 组合也是无穷无尽的。所以机器不可能通过统计解决“已知 N 个 Tokens 组合，求 M 个 Tokens 组合出现的概率”这个问题。

在多模态中，这个问题就更加突出。所以机器只能在有限数量的统计基础上，来推测“N 个 Tokens 组合出现后，M 个 Tokens 组合出现的概率”。这就是注意力机制的本质。预训练获得的权重矩阵就是有限数量的统计知识。而注意力机制就是基于有限数量的统计知识，来推出目前 Tokens，N 个 Tokens 组合后，其伴随 M 个 Tokens 出现的概率。如果在在 N+M 个 Tokens 中，有些 Tokens 权重高，就说明它们经常伴随出现，所以它们就更加可能是常见 Tokens 组合。

这就是注意力机制的核心机制，它就是一种寻找 Tokens 常见排列的方法。它的本质，可以认为是一种和神经网络结合在一起的贝叶斯推理。

所以注意力机制加持下的深度学习，所创建的坐标基底簇，更加符合人类创建概念的习惯。所以大模型和人类之间才可以实现语言交流[15]。

在语言模型中，“常见 Tokens 组合”就是“常用语”。它既包含常见 Tokens 的组织形式，这类似于语法的结构；也包含具体“常用语”；由于机器的统计分析能力远远超越人类，所以机器发现的“常用语（包含语法）”远比人类的“常用语”规模庞大的多。

注意力机制，非常类似于人类的学习。我们学习一本书中的信息时，“先读薄，再读厚”就是同样的方法。“先读薄”就是总结出其中的框架性信息，这是一个信息压缩过程；然后“再读厚”，就是在框架性信息的基础上，添加不同的细节（和其他矢量组合成新的矢量），来构成新知识，这就是一个信息生成过程[17][18][19]。

2.5 大模型的工作原理是什么？

在大模型中，当信息输入后，注意力机制的推理过程，就是把输入矢量向坐标基底簇的投影过程。注意力机制获得的权重，就是坐标值[15]。

在大模型中，输入 Tokens 在第一层向权重矩阵中矢量投影，这是一个矢量分解过程。然后，进行第二层投影，这就是输入 Tokens 组合并加权后，以组合的方式，再次向预训练权重矩阵的 Tokens 组合加权后，进行投影过程（组合到组合投影）。经过多层注意力机制操作后，就形成了多个输入 Tokens 组合到预训练 Tokens 组合的投影分解过程。

而最后一层注意力机制输出的权重系数矩阵，和它背后隐含的坐标基底簇（以常见 Tokens 组合作为坐标基底簇），共同形成了对输入信息的再描述（自注意力机制）。

所以，大模型的工作原理是：（1）它以预训练权重矩阵的 Tokens 组合为基底簇，而权重矩阵是通过试错法，从训练材料中获得的局部统计信息；（2）它采用注意力机制来实现输入 Tokens 组合向权重 Tokens 组合的投影过程（矢量分解），推理过程获得的权重就是坐标值。（3）有了矢量分量，就可以找到大量的临近矢量，这些临近矢量对应的下一个矢量，就是输出矢量。矢量的临近关系，以输出矢量的概率形式表现出来。

所以，大模型就是一个自回归预测模型。只不过，它在原始的输入基底上（每一个 Tokens 就是一个维度），进行了坐标基底簇转换过程。把“每一个 Tokens 就是一个维度”这样的原始基底簇，转换为“常见 Tokens 组合后，作为一个维

度”这样的坐标基底簇。然后进行自回归预测。

2.6 大模型为什么会有能力涌现？在什么时候涌现？

为什么大模型会有“涌现”现象？很简单的道理，比如一个美国人来到中国，他可以通过我们人类之间大量的共有背景信息（比如人身需求、社交结构等），通过中等数量的中英文对比，就能完成正确的翻译过程。

但大模型就像一个外星人，它和人类之间并没有共同的背景信息，它看到的东西，只有人类信息之间的连接方式。所以它需要提取人类信息之间的连接方式，来预测信息的发展过程。一开始，样本不够时，它提取的“信息框架”和人类“信息框架”差异很大，所以它会不断犯错误，在黑暗中摸索，总是四处碰壁。随着样本数量的不断增加，它的“信息框架”和人类“信息框架”有更高的概率对齐。但这不是一个线性过程。比如在提升到某一个阈值之前，它就像人类语言学家解密古代语言一样，在黑暗中摸索，进展甚微。在某一个节点上，如果正确率达到阈值，整个解密过程就会大大加快，急速完成。这就是“涌现”现象。机器“涌现”的并非智力，而是找到了正确的“常见的 Tokens 组合方式”。因为评价机器能力的标准是人类标准，所以当它的基底和人类基底接近时，它的能力就涌现出来了。

2.6 RLHF 能最终解决大模型面临的问题吗？

目前大模型存在两个严重问题：

(1) 幻觉问题[20]。

目前大模型的核心能力，是把输入信息转变到常见 Tokens 组合构成的坐标基底簇（矢量投影分解），这是一个信息空间的基底变换过程。

然后它利用获得的系数矩阵（注意力机制的推理权重），可以找到多个相似的“预训练矢量”（分量加权对比）。然后根据这些相似的“预训练矢量”，按照预训练获得的映射关系，找到“下一个矢量”，并选择其中一个输出。这就是自回归预测过程，也是 GPT 类大模型的工作原理。

所以，大模型优化的是“参数”。而每一个参数，背后对应的是一组 Tokens 组合。表面上，大模型在优化网络参数。其实质，是在优化常见 Tokens 组合，也就是说，在寻找一组最优基底坐标簇。神经网络的每一层系数，其背后都对应着一组隐含的基底坐标簇。

大模型从海量数据中获得的只有“常见 Tokens 组合”，并没有事实记忆。所以面对输入 Tokens，大模型只能通过分解输入信息到“坐标基底簇”上，然后获得不同概率下的下一个 Tokens。这个过程迭代进行，它本身就是一个创造过程。如果事实本身很“常见”，那么事实会以“常见 Tokens 组合”的形式被保留下来。如果事实没有以“常见 Tokens 组合”被保留下来，或者事实本身权重不够高，那么机器就会创造信息。GPT 本身就是信息生成，所以幻觉问题本来就是它本职工作的一部分[17][18][19]，所以这个问题，GPT 无解。

比如，机器发现很多记者的简介后面，都会有记者的其他文章网页链接，或者附上记者过去获得的奖项。如果机器见到这种信息组织模式很多，那么这种信息组织模式就会成为“框架”到“框架”的映射。所以如果输入信息中包含了类似的框架，但只是记者名字不一样，那么机器都可以通过“框架+细节”，映射到“框架+细节”，从而在输出也产生很多网页链接，或者是奖项。但这些网页链接和奖项也是通过“框架+其他矢量”映射到“框架+其他矢量”建立的，它们很可能根本就不存在！

为了解决大模型的幻觉问题，很多人指望外挂“向量数据库”，让大模型去

查询事实知识来消除幻觉。这是试图采用百科全书来实现通用人工智能的另外一个版本。无论是“向量数据库”，还是“知识图谱”，根本不可能解决幻觉问题！因为，这些知识是外挂的，和大模型自身的知识是无法融为一体的。它们就像一位普通人拿一本词典，就试图开一家翻译公司一样。当年专家系统碰到的问题，它都会碰到。

(2) 有害内容的问题[20]。

大模型中，注意力机制是对的，但深度学习有缺陷。

在大模型中，基于 Self- attention 的 Transform 模型，加入了位置编码，其主要目的是增加 Tokens 位置信息，使其可以利用每个元素相互之间的位置关系。这对注意力机制而言是必须的，因为注意力机制就是要找到 Tokens 的时间、空间关系。

但通过多层的深度学习网络，在误差约束下，大模型进行了多次坐标基底变换后，找到了“最优坐标基底簇”。但这种“最优坐标基底簇”的 Tokens 组合，和原始 Tokens 的时间、空间关系不再一样。虽然它可能依然保留有 Tokens 之间的部分组织信息（因为深度学习过程是不可逆的，所以 Tokens 的位置信息只会有部分被保留），但却难以为人类所理解和利用。所以，我们认为深度学习破坏了 Tokens 的原有时间/空间上的组织形式。

我们可以认为大模型执行了一次有损的翻译过程，把人类的 Tokens 组合次序，翻译为它的语言了。但问题是，人类并没有掌握大模型的语言，所以人类无法理解大模型创建的知识，也无法模仿其知识组织形式，给大模型植入“先天知识”，这就是问题的核心所在。

而且由于大模型无法实现小样本、累计学习，它需要超大样本，知识一次成形，这进一步增加了人类理解其知识组织形式的难度。

因为机器没有自身的需求，机器就不可能有自我感知的奖励和惩罚。机器没有自我感知的奖励和惩罚，就不可能自发创建矢量（信息）到奖励或者惩罚维度的投影。也就是说，机器所创建的基底坐标簇中，缺乏了奖励、惩罚、快乐、悲伤等人类特有的，也必须要有的基础维度！

目前大模型采用的补救方法是 RLHF。这相当于人类事后给特定矢量后面增加一个奖励维度的后缀。也就是说，机器的基底坐标簇中，增加了一个奖励维度。如果在训练数据中，在大量不同类型，足够数量矢量上，增加在奖励维度上的分量值，就相当于建立了这些训练矢量中的共有的分量组合，到奖励维度的投影。这就是机器的奖励函数。所以，机器也可以预测不同决策下，也就是按照不同的组合方式产生的输出矢量中，包含的奖励分量。所以，机器会优选奖励分量高的输出。这就是 RLHF 学习带来的惊人效果。因为通过 RLHF 学习的知识，实际上是可以泛化的。当一个机器有了自身的奖励、惩罚维度，就有了自己初步的“趋利避害意识”，这就是为什么我们会从目前大模型看到“意识”的朦胧影子。

但这是一种事后打补丁的方式，意味着需要机器先尝试，然后人类打分反馈，它只能用于可以大量试错的领域。这类似于一个孩子博士毕业了，但完全没有“是非”观念，父母只能跟在屁股后，喊“No”，“No”，“Yes”来赋予他“是非”观念，而且他和父母还无法直接交流，只能通过“Yes”和“No”来沟通。所以，这样的学习效果效率低，而且永远可能碰到那些意想不到的 corner case！

3 注意力机制+深度学习+强化学习，是通用人工智能的正确道路吗？

3.1 大模型就可以实现通用人工智能了吗？

我们认为，大模型证明了它的大方向是正确的。但我们并不认为大模型是实现通用人工智能的正确道路。

在 NLP 方面，人类从早期的词袋模型、词向量到 EMLO[21]，直到 Transformer，才真正地实现了注意力机制。把深度学习和注意力机制结合起来后[22]，就能产生类似于人类表达方式的优化的坐标基底簇，这就是 Transformer 能产生智力“涌现”的原因。

但我们注意到，大模型采用的道路是“先矢量化，建立初步关系；然后通过试错法，来调整坐标基底簇；然后在优选的坐标基底簇下，再次矢量化，获得正确的关系”。这样的机制，导致需要的数据量极大，计算量极大，并且知识是通过训练过程一次成型，难以实时更新[2][3]。

同时，奖励函数是在事后出现的，这对那些难以试错的领域，比如真实环境下的互动决策（自动驾驶、家庭保姆、工业、农业、商业、服务业、政府管理等），无法适用。

另外，“面向任务，搞强化学习”这种思想是错误的。人类之所以“通用”，是因为我们面对一切任务，都按照“趋利避害”来决策。机器也应该这样。任务千千万，面向任务搞强化学习，永远也学不完！而且很多任务试错成本很高！

3.2 什么样的道路，才是走向通用人工智能的正确道路？

目前大模型的问题是：

（1）注意力机制是对的。但深度学习有缺陷。

因为深度学习破坏了 Tokens 的原有时间/空间组织形式。导致产生的知识，难以被理解，无法被模仿。所以人类无法模仿其组织形式，给机器置入先天的“自我需求”（先天知识）。

机器没有“自我需求”，就不可能有“自己的想法”，就不可能“自主决策”。

这样，机器就只能按照预定流程（或预设，或统计），被动“决策”，无法灵活变通，这是目前 AI 的大问题。

（2）“面向任务，搞强化学习”这种思想是错误的。

人类之所以“通用”，是因为我们面对一切任务，都按照“趋利避害”来决策。机器也应该这样。任务千千万，面向任务搞强化学习，永远也学不完！而且很多任务试错成本很高！比如照顾孩子，没有人愿意把自己的孩子交给机器做实验！

所以，我们的解决方案是：

（1）既实现注意力机制，又不破坏 Tokens 原有的时间/空间组织形式。所创建的知识可以被理解，可以被模仿。

（2）我们可以模仿知识的组织形式，给机器赋予“先天需求”。“先天需求”作为一类特殊 Tokens，和其他 Tokens，通过注意力机制形成常见组合。这些常见组合就是常识（这就是世界模型）！

（3）机器只学一件事“如何满足自我需求”，也只处理一件事“如何满足自我需求”。这就是通用决策。

（4）因为没有破坏原有 Tokens 的时间/空间组织形式，所以机器可以通过语言符号直接获得 Tokens 的时间、空间排列方式。并且这种排列方式可以被理解，可以被模仿，所以机器可以通过语言学习，直接获得人类文明史上积累的所有经验！机器不再需要走一遍“进化史”！

4 实现通用人工智能的 Step by Step 步骤。

下面是实现我们方案的 10 个步骤。

Step 1，把信息 Tokens 化。（和其他 AI 技术一样）

Step 2，把 Tokens 矩阵化。（建立记忆库）

Step 3，输入 Tokens 按照相似性关系，向记忆库中 Tokens 传播激活值。

Step 4，所有被激活的 Tokens，按照临近关系，向临近的 Tokens 传播激活值。

Step 5，每一个被激活的 Tokens，又按照相似激活和临近激活原则，在记忆库中链式传播激活值。

其中，Step3~Step5 中，相似度越高，传递系数越大。存储位置越临近，传递系数越大。Tokens 的记忆值越高，传递系数越大。

Step 6，每个 Token 从不同传播路径获得的激活值，进行累计。

Step 7，所有 Tokens 的激活值，都随时间消退。

其中，Step3~Step7 是链式联想激活过程，这就是注意力机制的推理过程，激活值就是推理权重。

Step 8，每个 Token 按照其获得的激活值大小，按照正相关来更新记忆值。并且，所有的记忆值都按照时间而消退。

每个 Token 的记忆值就是它的预训练权重值。在记忆中，存在大量的 Tokens 组合方式，那些能重复出现的 Tokens 组合方式，它们包含的 Tokens 每次都能彼此激活，相互推高激活值，从而获得更高的记忆值。所以如果多个 Tokens 构成的组合出现在输入中，和这个组合相关的记忆中 Tokens 组合就有更高的概率获得高注意力权重。所以，链式联想激活过程是一个“Tokens 组合”优先的激活值传播过程。

Step 9，预置最小先天需求（先天知识，由 Tokens+记忆值+排列方式组成。）。先天需求，就是模仿知识的组织形式，建立的先天知识。先天知识可以包括最小的先天需求、奖罚、情绪和必要的先天安全本能知识，当然，也可以预置其他知识。这些知识是作为记忆库的一部分存在的，和后天形成的记忆无缝融合，形成整体记忆库。对先天知识的“Fine Tuning”是通过积累后天的知识（包含反馈）来实现的。

Step10，让先天需求、奖罚和情绪（使用特殊的 Tokens 来代表），和后天信息（普通 Tokens 信息流），在机器的训练中和生活中，形成时间信息流，并被存储。然后通过链式联想激活过程 + 注意力机制，形成全连接知识网络（记忆库）。

我们的方案，最后形成这样一个记忆库：每一个 Token，都是一条数据记录。它们由表 1 所示的 4 个字段构成。

表 1，每一条 Token 数据的组成。

字段 1	字段 2	字段 3	字段 4
时间标记	Token 可以是图形、语音或者其他传感器的数据	记忆值	激活值
代表 Tokens 彼此之间的时间关系	代表 Tokens 本身	代表预训练权重	代表注意力机制的推理权重

大量 Tokens 按照时间间隔存储起来，通过优化（通过链式联想激活过程+记忆和遗忘机制来优胜劣汰），就形成了知识网络。

知识网络，就是记忆库。其中的网络节点，就是 Tokens。其中的网络连线，就是激活值传递关系。但需要特别指出，激活值传递关系是由 Tokens 的相对位置、Tokens 的记忆值和 Tokens 之间的相似性，以及 Tokens 获得的初始激活值大小来决定的，所以是先有输入 Tokens 后，然后临时建立 Tokens 之间的激活值传递关系，这种传递关系并不是固定的。

其中的记忆值，就代表了预训练权重；其中的激活值，就代表了注意力机制下的推理权重。所以，在我们的方案中，知识获取和推理应用融为一体，先天知识和后天知识融为一体。

在记忆库中，既有客观 Tokens，又有主观 Tokens，它们通过注意力机制形成的连接关系就是“信息”。所有 Tokens 的排列关系就是全部信息，它的维度很高。而“知识”就是能够重复出现的排列方式（包括时间、空间），它们是信息中能够重复出现的那一部分，所以他们包含的 Tokens 更少，代表性更高，适用范围更大，更抽象，所以他们的维度更少。而常识”则进一步限定为我们人类常见的“知识”。

我们的机器，记忆库是可以置入、修改或者合并的，所以机器之间的知识是可以通过记忆库直接合并而共享的。比如，一个厨师机器人，通过载入医生机器人的记忆，就可以直接获得医生的各项技能。而不需要再次把“厨师大数据”和“医生大数据”合并后，花费数千万美金和几个月时间，重新做预训练。

4.1 每个步骤的详细说明。

Step1，把信息 Tokens 化。

机器只需要把输入信息打散，按照整体优先，按照低分辨率优先，提取其中的底层 Tokens（比如图像的整体轮廓，纹理，拓扑、线条，角、脊、顶点等，语音时域/频域音调、音色等主要底层 Tokens）。

按照时间顺序，依次存入记忆库就 OK。特别强调：不需要去识别它们，存下来就 OK。即使一开始提取的 Tokens 比较随机，算法不完善，也没有关系。因为我们这套算法，是通过不断积累的常见 Tokens 组合（也就是“世界模型”），在后续指导机器如何“按需提取”！常见 Tokens 组合，既包含常见 Tokens，又包含它们的组织形式。

所以机器提取 Tokens 这个过程，是一个逐步优化的过程。将 Tokens 存入记忆库后，随后按照链式联想激活过程，记忆和遗忘机制，不断改变这些 Tokens 的记忆值和激活值。通过优胜劣汰，那些广泛存在的 Tokens，或者 Tokens 组合会被保留下来，形成更加复杂的 Tokens。而那些很少能重复的 Tokens 则会被淘汰，它们不再被提取。

所以，机器处理 Tokens 的策略也是：寻找那些广泛存在的原始数据组合，作为 Tokens。这是常见信息组合优先原则在确定 Tokens 构成上的应用。这一点，类似于人类，它是进化带给人类的礼物。因为提取 Tokens 这样的底层程序，需要广泛的复用，才能达到能量的最大效用。

Step2，把 Tokens 矩阵化。

每一个 Token，对应记忆库中的一条记录，它有 4 个字段，如表 1 所示。记忆值大小表示记忆强度，为零则会被删除。激活值大小表示被激活的强度，为零表示没有被激活。所有记录按照同时性存储方法，就自发构成了整个记忆库。

关于同时性存储方法，具体实施方式包括：

（2.1）机器保留 Tokens 出现在输入信息中的时间相对位置。

一种实现方法是：机器使用 Tokens 在存储空间中的距离来反映这些 Tokens 被存储的时刻之间的时间距离，比如机器按照输入的时间次序来依次存储 Tokens，时间越临近的 Tokens，存储位置越临近；

另外一种保留时间相对位置的存储方法是每个 Tokens 都带有记忆空间中的坐标；记忆空间中的坐标，主要包括 Tokens 的存储时间信息；

机器保留 Tokens 出现在输入信息中的空间相对位置；一种实现方法是：机器把每一次提取的 Tokens，按照和原始数据相似度最高的位置、角度和大小，把它们和原始数据重叠放置，并在存储时保留这些 Tokens 在空间上的相对位置；

实现方法还可以是：整体低分辨率 Tokens 优先提取，然后根据机器的决策，再按需提取其他局部 Tokens。这样，通过临近存储关系，局部 Tokens 和整体 Tokens 既存在临近激活关系，又存在 Tokens 之间的相似性关系，所以它们会彼此激活，建立位置关系连接。

Step3，从输入 Tokens 到记忆库中 Tokens，进行相似性激活。

给输入的每个 Token 赋予一个统一的初始激活值 A0。A0 本身是一个预设的数值。但它可以受到上一次机器链式联想激活过程中，被激活的奖励符号、惩罚符号的激活值高低进行调整。

被激活的奖励符号、惩罚符号的激活值高低，就是机器对之前输入信息进行的潜在奖罚值进行预测。而初始激活值 A0，会影响链式联想激活过程的范围。当初始激活值 A0 很高，那么链式联想激活过程的传播范围就更大。这是因为在我们的方案中，激活值传播系数是小于 1 的。随着链式传播的级数增加，被传播的激活值越来越小。当一个 Token 获得的激活值小于预设的阈值后，链式传播过程就会终止。所以 A0 反映了机器对输入信息的重视程度。当 A0 很高时，机器会激活更多记忆中的 Tokens，来寻找和输入 Tokens 相关的记忆。这和人类类似，如果前面输入的 Tokens 带来了很高的潜在奖罚，那么新的相关 Tokens 输入就可能被格外重视。比如，老板的话，会让你联想的信息更多。

相似性激活的原则是：（1）Tokens 之间相似度越高，传递系数越大；这是 Token 之间的相关性点积。（2）记忆值越高，传递系数越大；记忆值是预训练权重。需要强调，同一 Token，可能在记忆库中很多位置上不断出现！它们都有自己的不同记忆值！这是因为不同 Token 排列下，同一 Token 在其中的权重并不相同！这和大模型中注意力机制是类似的。

Step4，所有被激活的 Tokens，进行临近性激活。

我们认为，Tokens 之间的临近关系，代表了它们之间存在某种隐含的关联性。出现时间上越临近，潜在关系越紧密。这种关联性可以通过链式联想激活过程 + 记忆和遗忘机制统计出来。临近关系，实际上反映了一种 Tokens 组合关系。如果这种组合关系能重复出现，那么它就是一种常见组合。所以我们通过链式联想激活过程中的临近激活过程，来发现常见组合方式。

每个被激活的 Token，又会向它临近的 Token 传递激活值；时间位置越近，传递系数越大；记忆值越高，传递系数越大。记忆库中，Tokens 之间如果存在临近关系，说明了它们曾经是一种组合方式。如果它们的记忆值高，说明了它们是一种常见的组合方式。如果只有一个 Token 的记忆值高，说明它们不是常见组合方式。如果 Tokens 的记忆值都不高，则它们传播的激活值很低，Tokens 的链式传播很快停止。说明这样的信息不重要，它们在信息处理中的权重很低。

Token 采用“时间位置越近，传递系数越大；记忆值越高，传递系数越大”

的方式，激活包含它们的常见组合，本质就是输入 Token 向一组 Tokens 组成的坐标基底的投影过程。

如果 N 个输入 Tokens，它们都向记忆中包含它们的 X 组合（Tokens 组合）投影，那么这些 X 组合就会获得很高的激活值。因为每一个 Tokens 都会同时按照相似性和临近性激活 X 组合中多个 Tokens。所以，X 组合通过激活值累计的方式，获得了更高的激活值。这些更高的激活值 Tokens，组成的“模型”，就是输入的矢量（N 个 Tokens）激活的预期模型（世界模型）。

本质上，这就是一个矢量向坐标基底分解的过程，也是信息识别过程。

Step5，每一个被激活的 Tokens，又按照相似激活和临近激活原则，在记忆库中链式传播激活值。

每一个输入的 Token，都在记忆库中进行“相似性激活”、“临近性激活”，激活值传递大小和它们的预训练权重（记忆值）正相关。

记忆库中每一个被激活的 Token，同样按照“相似性激活”、“临近性激活”，激活值传递大小和它们的预训练权重（记忆值）正相关。

这个过程链式进行，直到所有的输入 Token 完成自己的“链式激活过程”。所以，除了和输入矢量相似的记忆中矢量被激活外，机器还会激活和输入矢量相似的记忆中矢量的“前因”和“后果”，也就是在记忆库中，在时间上的前面信息和后面信息。并且，可能通过不同的记忆片段，激活不同的“前因”和“后果”。这就使得我们的方案能推测可能的前一个矢量，并预测可能的下一个矢量。

由于我们采用的策略是“整体低分率 Tokens”优先，所以信息的空间位置关系，实际上是通过时间位置关系来建立的。当信息输入时，机器首先提取的是“整体低分率 Tokens”，存储到记忆库中。随后发起链式联想激活过程。完成后，通过统计被激活的奖罚符号的激活值，来做决策。

机器的决策原则是趋利避害。做出的决策有可能是进一步识别信息，或者其他决策。如果决策是进一步识别信息，那么机器会把目前的高激活值 Tokens 组合方式（包含语言 Tokens）作为预期模型，去主动确认那些还没有出现在输入中的高激活值 Tokens。采用的方法是模仿过去获得这些 Tokens 的经验，来调整自己的传感器系统。所以这是一种主动寻找信息的“模式识别”，和人类的识别过程是类似的。

这些新获得的 Tokens（比如局部细节），就和原来的“整体低分率 Tokens”存在时间上的临近性关系，也存在部分相似性关系，所以它们之间可以通过彼此传递激活值来建立连接关系。这样，新获得的 Tokens 就和原来的整体低分率 Tokens 建立了位置关系。这些整体低分率 Tokens，和那些经常伴随出现的局部 Tokens，通过记忆和遗忘机制，慢慢就形成了“世界模型”。

需要指出，世界模型并不是创建一个独立的模型，它所包含的 Tokens 可能遍布在整个记忆库中，这些 Tokens 是通过相似性、临近性和高记忆值带来的紧密激活值传递关系而临时创建的。所以它不是静态的，是分布式存在的，是在输入信息激励下，那些获得高激活值的 Tokens 临时构成的，记忆库中并没有单独的模型存在。

Step6，激活值累计。

如果某一个记忆库中的 Token，和多个输入 Tokens 之间存在激活值传播路径（也就是说，要么直接相关，要么间接相关），从输入传递过来的激活值是累计的。所以和多个输入 Token 之间存在直接/间接相关的记忆库中的 Token，会从多个传播路径上，获得更高的累积激活值。

通过这种方式，输入 Token 中，如果彼此存在关联的 Tokens，会彼此推高记忆库中的相关 Tokens 的权重。也就是说，那些常见组合，它们的激活值，会从激活值海平面上升起来。而这个激活值海平面就是那些大量 Tokens 的低激活值。那些从激活值海平面上升起来的 Tokens，就构成了一个或者多个“世界模型”。

而那些和输入最相关的记忆，尽管它们可能不是常见的，但由于和输入直接相关，传播路径短，所以它们可能也能获得高激活值。

所以，我们的方案，既能通过常见 Tokens 组合来获得信息的“信息框架”，又能关注特定事实细节，所以我们的方案，是自带“事实数据库”的，它能解决目前 GPT 的“幻觉”问题。

Step7，激活值随时间消退。

所有的激活值，都随时间而不断递减。当后面的 Token 输入后，激活了记忆中相关 Token。而前面输入，所激活了的相关 Token 还没有完全消退，激活值会被累计。

而机器的决策，是基于所有被激活 Tokens 的。所以前、后输入信息都会被考虑到。所以，机器的思维是有一定时间连贯性的，可以解决“省略”、“代指”、“比喻”等问题。

所以，我们的机器，利用了前、后输入之间的隐含关系！这就是注意力机制！

更进一步：机器会根据上一次决策所预测的“利弊”大小，来调整给输入 Tokens 赋予的初始激活值 A0。而初始激活值 A0，会影响激活值传播的范围和累计的大小！这就是根据“利弊”来调整注意力强度！这和人类非常类似。在这一点上，超越目前技术（Transformer）。事实上，这和人类的决策过程很相似，比如老板的话，会能让你产生更多的联想，激活更多的奖励或者惩罚符号，从而更深入的预测奖罚值。

Step8, 通过链式联想激活过程 + 记忆和遗忘机制 + 趋利避害原则来更新预训练权重矩阵。

在我们的方案中，那些能够重复出现的 Tokens 组合，因为重复性，它们可能获得更高的记忆值。并因为是能重复出现的组合，每一次彼此都推高对方的激活值，所以获得了远比简单的重复性更高的记忆值。

而且因为它们能重复，所以它们的组合，每一次都能获得更高的激活值，所以它们更加容易被激活，从而更加容易获得记忆增量。所以这是一个正向循环过程。所以，从这里可以看出，我们的机器可以自我总结经验。但同时，要遗忘已有的思维模式，也会是一个费时的过程。

所以，在我们的方案中，机器的预训练统计过程，并不是简单地统计重复性，然后采用记忆和遗忘机制来建立的。而是通过注意力机制 + 记忆和遗忘机制+趋利避害原则，来共同完成的。

机器的决策过程，是趋利避害的，机器在趋利避害的决策中，对信息的识别过程，是根据趋利避害的方式，对信息做选择性识别的。所以，我们的机器，是根据自身的需求，来建立 Tokens 之间的常见组合的。所以，我们的机器，对外界、对自身的信息识别，都是选择性识别的。

记忆和遗忘机制：记忆库中所有 Tokens，如果被激活一次，就按照它们的激活值大小，正相关更新它们的记忆值。它们的记忆值，就是预训练权重矩阵！由于 Token 排列无法穷举，所以这是一种非完全统计过程，和大模型的预训练过程是类似的。

而链式联想激活过程，就是在输入 Token 组合激励下，注意力机制的推理过程（从局部统计权重到输入的本地化权重计算过程），这和 Transformer 中的注意力机制是相似的。

这个过程，本质上就是输入矢量向注意力机制建立的坐标基底簇投影的过程。输入矢量，可以看做是输入维度下的脉冲函数构成的原始基底簇。而注意力机制建立的坐标基底簇，则是以常见 Tokens 组合为基础建立的。

注意力机制的推理权重矩阵就是输入矢量到基底簇投影的系数矩阵。而在我们的方案中，链式联想激活过程，和 Transformer 中的多层注意力机制类似，也是输入矢量向注意力机制建立的坐标基底簇投影的过程：先单独的 Tokens 投影，然后组合投影。最后链式激活完成后，高激活值 Tokens 组成的一个或者多个高权重分量，就是信息的“框架”。每个框架包含很多 Tokens，难以具体描述。但通常其中的语言符号，由于代表性高，重复性高，所以通过获得的激活值可能也是最高的，它们就可能成为这个“框架”的代表性 Tokens。所以，我们方案中，激活值就是推理权重矩阵。

事实上，无论是大模型，还是我们的网络，都是一种类神经网络。注意力机制，本质是贝叶斯推理。通俗的说，注意力机制，就是已知一些 Tokens 的条件概率，和部分 Tokens 的联合概率，求特定 Tokens 组合的联合条件概率。这就是贝叶斯推理和神经网络结合起来的应用。在大模型中，已知一些 Tokens 的概率，和部分 Tokens 的联合概率是由权重矩阵确定的，并通过多次相关运算来进行 Tokens 组合下的概率预测。在我们的方案中，已知一些 Tokens 的概率和一些 Tokens 的联合概率，是显式地被表达在记忆库中，它们就是 Tokens 的记忆值，Tokens 的相对位置和 Tokens 之间的相似性。

可以看到，我们实现注意力机制的方式是小样本、累计学习的。而且权重矩阵是实时更新的，所以我们的方案，知识是实时更新的。而其我们并不区分预训练和推理过程，所以我们的机器是终身学习的。

另外，可以看到，我们的方案，不需要 BP 算法，不需要预训练，它的运算量基本和大模型的推理过程接近。所以我们方案需要的计算量远小于 Transformer，并且同样可以并行计算。所以，我们的方案，可以实现预训练过程的计算本地化。每一个机器，都是一个自我训练，不断迭代，不断进化的智能体。

另外，可以看到，我们方案中，Tokens 提取和目前大模型可以采用类似的技术，运算量是相当的。而链式联想激活过程，是高度模式化的，它可以采用新型存储器件在硬件层面直接实现。这样，有助于我们方案中计算的本地化，这将有助于拓展落地场景，并降低成本。

Step 9，预置最小先天需求。

我们既实现了注意力机制，找到了常见 Tokens 组合，又没有打乱原有的 Tokens 的时间、空间组织形式！所以，我们方案形成的知识网络，是人类可以理解的。所以，我们可以模仿最终记忆库中 Tokens 的组织形式，给机器建立最初的最小先天记忆！这就等同于给机器预置一段类似于人类的最小先天知识（婴儿天生就有的知识）。

在先天记忆中，需要包含机器的最小“需求系统”、“奖罚系统”和“情绪系统”。采用的方法是：使用特殊 Tokens 来代表每一种“需求”、“奖罚”和“情绪”。然后模仿记忆库预训练后的形式（其实就是合适的 Tokens 排列方式 + 合适的记忆值），植入最小先天知识。

在日常生活中，让这些代表“需求”、“奖罚”和“情绪”的 Tokens 和其他引发它

们的外界 Tokens 一起训练，一起链式联想激活，一起记忆和遗忘。也就是说，通过注意力机制，让这些特殊 Tokens，和其他 Tokens 一样，建立常见 Tokens 组合。所以，我们必须要预置机器的最小"需求系统"、"奖罚系统"和"情绪系统"，这样才可能让代表外界(包括机器自身状态参数)的 Tokens，引发这些特殊 Tokens，从而建立起信息流。并通过链式联想激活过程 + 趋利避害决策 + 记忆和遗忘机制，来逐步获得最常见的、和机器最关心的常见 Tokens 组合。

这样，我们就在"客观世界的常见 Tokens 组合"和"需求"之间建立了连接关系。"客观世界的常见 Tokens 组合"就是客观世界的"客观常识"，而"客观世界的常见 Tokens 组合"和"需求"构成的"常见 Tokens 组合"，就是"主观常识"。"客观常识"和"主观常识"构成了"常识"。

常识就是"世界模型"，它包含了人类对外部世界认知的"世界模型"，也包含人类建立的"世界模型"和"我"的关系。需要特别指出，Tokens 不仅仅是静态特征，也包含那些简单的动态特征（比如旋转、摇摆等），所以世界模式不是静态的，也不是固定的，是在输入 Tokens 激励下临近创建的！

而且每一个人所建立的世界模型都是不一样的，这和它的经历直接相关。在我们的方案中，机器所建立的"世界模型"直接和它的训练数据相关，也会和它的生活经历相关！

有了世界模型，输入 Tokens 就能通过链式联想激活过程，去激活那些奖罚 Tokens、情绪 Tokens、需求 Tokens，而从输入 Tokens 到这些特征 Tokens 的激活值传递路径，就是和神经网络兼容的逻辑推理过程！它是显式的，是可以被理解的，可以被模仿的，所以机器的决策是可以看到的。

事实上，在实际创建"通用人工智能"过程中，Step 9 本质上是第一步。但我们可以通过前面的步骤来训练实验数据，从而获得并理解机器创建的知识的组织形式，然后模仿这些组织形式，来实现 Step 9。

① 预置和机器生命活动相关的，基础需求利弊系统。

比如给电量数据一个合理区间，在"先天记忆"中预置一个代表"饿"的符号，在"饿"符旁边放一个"惩罚"符号和一个代表"饿"的情绪符号。并赋予它们合适的记忆值。

当电量不够时，生命状态监控程序，会直接给"先天记忆"中"饿"的符号赋予初始激活值。它的激活值就会在整个记忆库中链式传播。它旁边的"饿"的情绪符号被激活，它旁边的"惩罚"符号也会比激活。所以机器就有了"饿"的情绪和出现"惩罚值"。为了避免"惩罚值"，机器会利用自己的经验，主动去寻找插头充电！

② 预置机器价值观的"高阶需求"利弊，需要预置最简单的沟通手段，然后培养价值观。

价值观需要从小培养！所以我们需要从小通过教育，来培养机器人的"价值观"。既然要教育，就需要通过"奖励"和"惩罚"来实现。所以机器一开始，就需要能够识别"奖励"和"惩罚"。这样我们才能通过"奖励"和"惩罚"，来发起第一步的学习！

所以，我们需要模仿后天记忆网络组织形式，让机器拥有能够识别最简单"奖励"和"惩罚"的先天知识！

比如：预置最基础的点头特征（假设 X 个 Tokens） / 摇头特征（假设 Y 个特征），不需要精确！

在点头 Tokens 旁边，放一个"被尊重"符号；在被"被尊重"符号旁边，放

一个“奖励”符号；给这些符号，赋予较高的记忆值，让它们之间的关系，成为长期记忆。当信息输入中，出现部分点头 Tokens 时，通过链式联想激活过程，机器就获得了“奖励值”。为了追求“奖励值”，机器以后可能会规划出各种决策，目的就是获得“人类的点头”！

类似于一个孩子，从最简单的沟通方式开始，逐步获得复杂学习能力，他（她）逐步建立的“奖励函数”逻辑链是：“奶”→“奶嘴”→“奶瓶”→“奶粉罐子”...→ ....“学习成绩”→“房子车子”... →“社会地位”....→“人生理想”。

所以，经过训练，机器的记忆库中，存在大量的奖罚相关的 Tokens 符号，和与这些奖罚 Tokens 关系密切的 Tokens 组合，它们之间存在因果关系。这些和奖罚 Tokens 关系密切的 Tokens 组合，它们代表的事物、行为和结果，就是价值观。所以，机器的任何价值观，都可以通过预置先天的沟通手段，然后一步步进行培养而建立起来。事实上，人类也是这样的，没有人先天就是“圣人”。

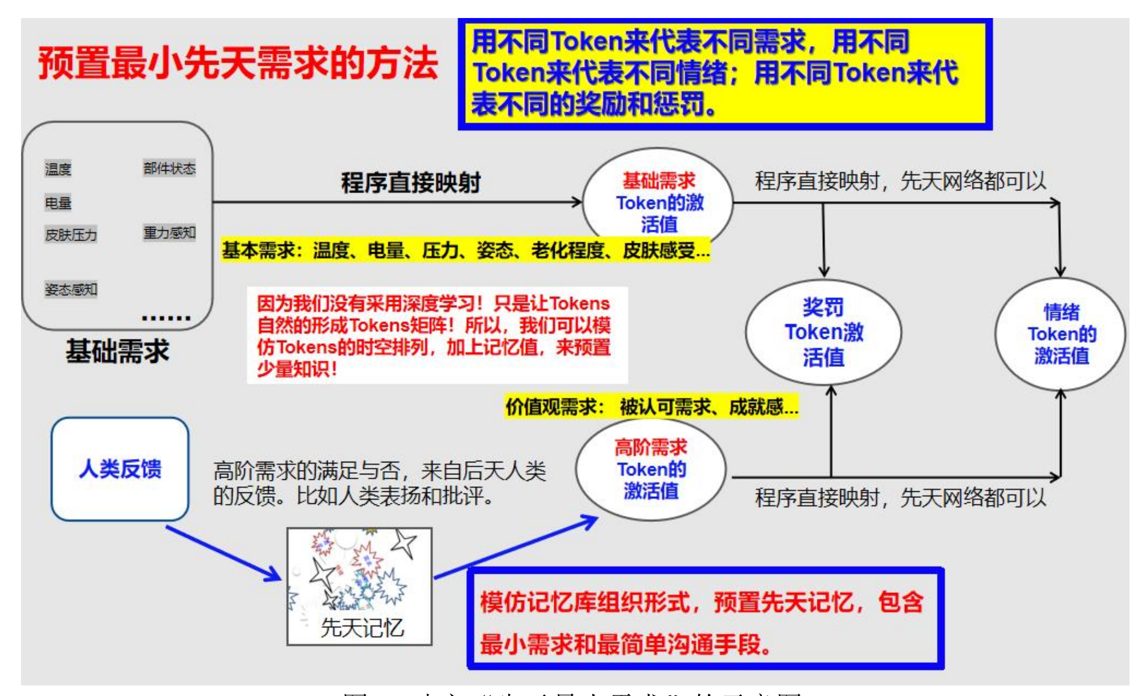


图 1 建立“先天最小需求”的示意图。

Step 10，形成全连接知识网络。

我们的方案，最后形成这样一个网络：每一个 Token，都由 4 个字段构成：它们分别是时间标记、Tokens 本身，记忆值和激活值。

大量 Tokens 按照时间间隔存储起来，通过优化（采用的方式是：链式联想激活过程+记忆和遗忘机制来优胜劣汰），就形成了知识网络，其中的记忆值，就代表了预训练权重；其中的激活值，就代表了注意力机制下的推理权重。

我们的网络，既有客观 Tokens，又有主观 Tokens，它们通过注意力机制形成的连接关系就是知识，其中常见知识就是“常识”。

这就是我们的机器，能预判利弊，能自主决策的原因！因为它有“需求”，还有和“需求”相关的“逻辑链”（Tokens 构成的激活值传递链路）。在需求的驱使下，它会主动去学习，自我迭代！比如自己去充电，自己去找图书馆看书！

在我们的方案中，知识是围绕“需求”展开的，决策也是围绕“需求”展开的， 这就是我们的机器能实现“通用”的核心原因！ 它面对的只有一个任务：

“需求”，而不是形形色色的“外界任务”。所以，我们的方案是“主动型智慧”，而目前所有其他方案就是“被动型”智慧。

可以看到，我们的方案是小样本学习、知识实时更新，训练和使用过程是一体的，所以机器是终身学习，自我迭代。

由于机器的知识是以记忆库的形式存在的，而记忆库又是按照时间次序存储起来的，只不过在原始记忆库的基础上，逐步优化了记忆值。所以不同的记忆库可以直接拼接起来，形成大的记忆库。所以，厨师的记忆库和医生的记忆库，融合后，机器人就能同时拥有厨师和医生的技能，而不需要把大量的厨师和医生的数据放在一起重新训练。而目前的 AI 技术路线，则无法实现这一点。在大模型中，必须要同时使用大量的医生和厨师数据进行训练，机器才可能同时掌握这两种技能。显然，按照这样的训练方式，希望机器能拥有“各种各样”的能力是一种奢望。

4.2 记忆值和激活值变化过程的一个示例。

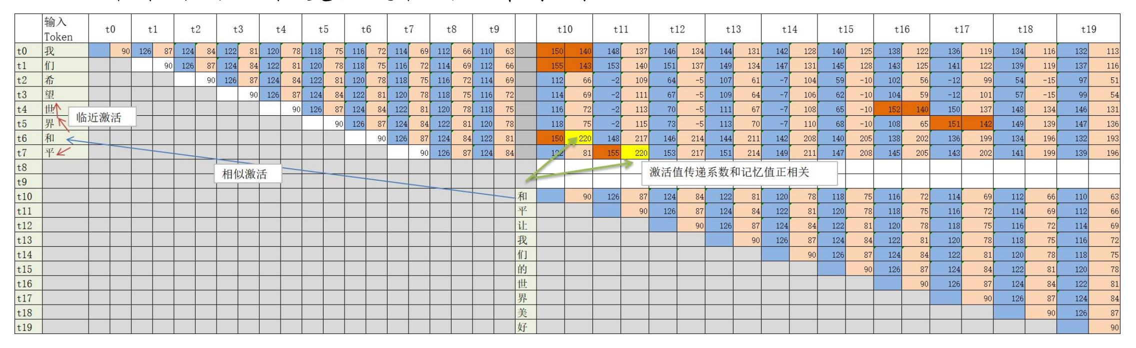


图 2，联想激活过程中，记忆值和激活值的变化过程简单示例

图 2 为联想激活过程中，记忆值和激活值的变化过程的简单示例。为了简化，假设这时机器的记忆库是空的，机器是“机生”第一次接收输入信息（而且我们也没有给机器预置先天记忆）。假设，在 t0 到 t7 时刻，机器的输入 Tokens 是“我们希望世界和平”。在实际流程中，机器应该根据目前被激活的奖罚符号的激活值大小（价值预估），来调整给所有输入 Tokens 的初始激活值。但在这里，由于没有价值体系来调整，我们假设默认赋予给输入 Tokens 的初始激活值为 90（假设激活值区间是 0~255），所以链式联想激活过程后，假设按照记忆曲线，Tokens 的激活值是 90，而目前记忆值为 0 的情况下，机器获得的记忆值增量为 126。

记忆值更新增量$\delta m = f(m0, A0)$，其中 m0 代表目前记忆值，而 A0 代表目前激活值。记忆值更新增量和激活值成正相关。

所有的记忆值和激活值都随时间而递减。这里采用了夸张的递减梯度。

在 t9 到 t19 的时刻，机器接收到第二次输入 Tokens：“和平让我们的世界美好”。显然，按照“相似性激活”过程，首先 Token“和”会激活记忆库中的 Token“和”，给它传递激活值，并且由于记忆库中“和”的记忆值较高，所以记忆库中的“和”从输入 Tokens 的初始激活值，获得了传递过去的激活值，而且这个传递系数较大。

相似性激活过程传递系数 T= f(S , m0)，其中 S 代表相似性（Tokens 矢量的点积），m0 代表被传递的 Tokens 的记忆值。激活值传递系数和相似度、记忆值正相关。

同时，在记忆库中的“和”Token，还会因为激活值超过预设阈值，而发起链式传播过程。在这个链式传播过程中，它首先会通过“临近激活”方式，向和它临近的“平”、“界”发起临近关系激活。

临近激活过程传递系数 T= f(D,m0)，其中 D 代表两个 Tokens 的时间距离，m0 代表被传递的 Tokens 的记忆值。临近激活值传递系数和时间距离成反相关，和被传递的 Tokens 的记忆值成正相关。

而“平”、“界”获得了激活值后，如果激活值超过预设阈值，也会发起链式传播过程。在记忆库中寻找和自己相似的 Tokens 进行激活值传播，也会对和自己临近的 Tokens 进行激活值传播，两个过程的传递系数都和记忆值成正相关。

通过链式联想激活过程，输入 Tokens，有可能激活整个记忆库和它们相关的 Tokens 组合。激活范围取决于它们获得的初始激活值，初始激活值受价值预测的调整。

在第二次输入的所有 Tokens 完成链式联想激活过程后，我们可以看到在记忆库中存储的 Tokens 中，其中“和平”Tokens 组合的记忆值最高，并且临近，所以它们在以后的链式联想激活过程中，每一个 Token 都会因为高记忆值而获得更高的激活值。同时，“和”、“平”除了自己有机会获得更高的激活值外，它们还会因为位置临近而彼此传递激活值（临近激活），而且这个过程中，同样因为它们的记忆值高而获得高的传递系数，通过激活值累积，它们就是一组容易获得高激活值权重的 Tokens 组合。

其次，我们可以看到在记忆库中存储的 Tokens 中，其中“世界”Tokens 组合的和“和平”Tokens 组合类似，获得第二高的记忆值，所以它们也是容易获得高激活值权重的 Tokens 组合。

所以，我们只需要两句话，就能建立 Tokens 的相对“权重”。按照上述过程不断累积学习，机器就能建立起正确的常见 Tokens 组合，以及它们的记忆值。而这种记忆值就对应了这种组合的“常见程度”，也就是说，记忆值其实就是通过训练数据获得的这种组合出现概率的局部统计值。而激活值，则是依据这种局部统计值，来获得的输入 Tokens 组合到“常见 Tokens 组合”（基底簇）的点积过程（投影）。

所以，我们采用的是一种类人的学习方法，它非常高效，并且可以实现小样本、累积学习、实时更新。它不修改“旧知识”的参数，所以不会有“灾难性遗忘”的问题。它不需要 BP 梯度优化过程，所以它的计算量基本和大模型的推理过程一致。

5，我们实现了 Yann Lecun 教授提出的三个条件。

深度学习三巨头，图灵奖获得者，Yann LeCun 教授认为 AGI 正确的方向是“ 世界模型”，道路是实现“类人 AI”，他们提出 3 个条件：

（1）：需要世界模型。包括需要要有对快乐、饥饿等基本需求进行建模的需求模块，以及预测价值的价值模块。

（2）：需要一种和神经网络兼容的逻辑推理能力。（目前推理能力都是靠外挂的符号主义推理）。

（3）：需要一种的“通用决策能力”，能自顶而下，分解决策。而不能对每一种任务都去强化训练 100 万次！

他们虽然提出了这些思想，但没有完整的技术方案。而我们的方案，可以实现上述 3 个条件。

5.1 我们建立了世界模型。

输入 Tokens 激活了记忆中 Tokens 组合，高激活值 Tokens 组合就是被激活的

"世界模型"（一部分 Tokens 可能已经出现在输入中，其他 Tokens 可能还没有出现在输入中）。然后根据预测的"趋利避害"的决策流程，决定要不要进一步确认其他"高激活值 Tokens"是否存在，这就是"模式识别"。世界模型就是"常识"，它就是"需求"、"奖罚"和"情绪"等主观 Tokens 和客观 Tokens 构成的 Tokens 组合方式。人类就是用"常识"来对事物进行"模式识别"的。

机器在每一次新的信息输入后，都需要进行链式联想激活，然后按照"同时性存储"方式存储 Tokens。同时性存储是指采用某种机制，来反映 Tokens 之间的时间间隔关系。比如可以按照时间越临近的 Tokens，存储位置越临近，或者按照每个 Tokens 所带的时间信息来确定时间间隔。

每一次获得新 Tokens 后，机器都需要更加更新后的激活值，寻找实现奖励、避免惩罚的路径。这些路径的集合就是整体响应路径。整体响应路径可能是一种网络状结构，很多局部路径既可能通向奖励符号，也可能通向惩罚符号。

由于有了通向奖励符号（或者惩罚符号）的激活值传递路径，也就是说，我们实现了奖罚函数的前置化和步骤化。所以，我们就解决了目前强化学习过程中，奖励函数稀疏和滞后的问题。机器通过类似于 AlphaGo 的最优响应路径搜索过程，就可以找到初始的最优响应路径。

如果整体的奖罚值累计没有进入可以接受的预设值（或者没有收敛），机器无法决定是否选用或者排除某些特定的路径，从而达到利益最大化。则机器需要进一步识别输入信息，增加更多的 Tokens，来对某些特定的奖罚激活值传递路径进行细分，从而进一步帮助机器选用或者排除某些特定的路径。这一步就是机器自发创建的、主动寻找信息来帮助决策的过程。这个过程迭代进行，直到奖罚值统计达到接受的预设值或者收敛为止。

在进一步识别输入信息时，高激活值 Tokens，要么是因为它们的记忆值高，比如是一类事物的代表 Tokens，要么是和本次输入 Tokens 关系紧密的 Tokens，比如相似，或者经常临近出现。所以记忆中被激活的高激活值 Tokens 组合，就是和本次输入信息相关的代表性 Tokens 组合，这些代表性 Tokens 组合，就是机器临时创建的"世界模型"，我们称之为"预期模型"。它既来自于过去经验的总结（优胜劣汰后的 Tokens 记忆值），也和目前具体输入直接相关。它是通过高激活值临时创建的，是机器对目前输入 Tokens 组合的"预期模型"。

机器参考"预期模型"中已经在输入中出现的 Tokens 和没有在输入中出现的 Tokens 之间的空间或者时间关系，以目前已经出现 Tokens 的时间和空间位置为基准，预测那些还没有出现的 Tokens 可能出现的时间或者空间位置;这些在预期模型中还没有出现的高激活值 Tokens，就是预期 Tokens; 机器按照预期 Tokens 在预期模型中的时间、空间和大小来分配机器的传感器搜索的时间和空间位置，并根据预期 Tokens 的属性（比如语音、图像或者触觉）来确定采用的传感器类型，并根据预期 Tokens 的属性（比如大小）来确定需要使用的分辨率。这就是机器的"按需识别"过程。这个过程可以迭代进行。

选择性注意力用于从输入信息中提取 Tokens 的一种手段，机器按照选择性注意力识别给出的识别区间和分辨率，从输入信息中提取 Tokens。这样才能解决图像信息的无线粒度化问题（机器按需提取图像中的信息）。机器在中提取特定区间数据时，按照整体特征优先的方式，优先提取选定区间内整体拓扑、外形轮廓、主要线条和主要纹理等 Tokens。然后，机器通过链式联想激活过程，在记忆网络中获得相关的记忆，并把这些记忆按照权重高低组合成不同权重的预期模型。

机器根据被激活的奖罚 Tokens（奖罚 Tokens 的激活值大小，就是预期的奖罚

值大小），使用决策过程，来决定是否进一步识别输入信息，还是对输入信息做出响应。

如果机器决定进一步识别输入信息，机器通过模仿过去获得“预期 Tokens”的相关经验，来进一步从输入信息中提取“预期 Tokens”。因此，机器是通过注意力机制，不断迭代提取输入信息的 Tokens，而每一次提取过程，可能使用不同的传感器，针对不同的识别区间，采用不同的分辨率。所以同一输入事物，机器可能提取到不同类型、不同区间和不同分辨率的 Tokens，并使用这些 Tokens 组合来构成同一事物的“分层表征”。“分层表征”是指按照区间内低分辨率的整体特征优先的方式，来逐次提取信息的 Tokens。

采用高激活值 Tokens 来构成预期模型;它的理论基础就是这些高激活值 Tokens 来自于两个部分：一是同类事物的共有特征；因为共有特征广泛存在于同类事物中，所以它们的重复性很高，所以它们通常是高记忆值 Tokens。所以在我们的方案中，机器对信息的识别方法是，首先通过共有特征来识别大的类别（获得抽象概念），然后才是通过迭代方法，逐步加入更多的 Tokens 来限定范围（从抽象概念走向具体概念）。

高激活值的另外一个来源是：在输入的 Tokens 中有和特定记忆中相似的 Tokens。这些特定的 Tokens，会因为相似性激活而被直接激活记忆中的 Tokens，和它存在临近关系的其他高记忆值 Tokens 也容易获得更高的激活值。由于激活路径短，所以在关系网络中，特殊 Tokens 会激活特定“预期模型”，这是一种通过特殊 Tokens 快速定位预期模型的途径。

所以对输入信息的识别过程，是通过共有特征识别其属于哪个大的类别，然后通过独有特征去确定其属于哪个具体的子类。机器通过选择性注意力，不断迭代增加用于识别的 Tokens。在这个过程中，先前被激活的 Tokens，其激活值会随时间而消退。如果它们被新输入的 Tokens 再次激活，它们的激活值会持续保持。如果它们和新输入的 Tokens 无关，则它们的激活值慢慢消退，逐步退出决策过程。

“世界模型”包含两个方面：1，机器认识世界是按照“模式识别”的方式来迭代进行的。2，机器是按照“利弊价值”的方式，来认识世界的。这是因为“利弊价值”是人类建立的核心“世界模型”。它是指导人类一切行为的“世界模型”。

所以，我们实现了“世界模型”。

5.2 我们实现了和神经网络兼容的逻辑推理能力。

所有被输入 Tokens 激励的“奖罚”Tokens，它们的激活值大小就是价值预测。

从输入 Tokens 到被激活的奖罚 Tokens 的传播路径，就是和连接主义完全兼容的推理能力！记忆网络是由 Tokens 按照激活值传递关系组织起来的神经网络。激活值传递的本质，就是实现注意力机制的推理过程。

Tokens 组合中的每一个 Token，通过链式联想激活过程，激活了和自己常见的 Tokens 组合，通过激活值累计过程，就能实现从输入组合（已知 N 个 Tokens 的特定组合概率），求出和输入最相关的 Tokens 组合（求 M 个 Tokens 的特定组合概率），而记忆库中最终的激活值分布就是获得的贝叶斯推理结果。

事实上，目前大模型中的注意力机制，已经实现了和神经网络兼容的逻辑推理能力。但存在两个缺陷：1，深度学习破坏了原有的 Tokens 时间和空间的组织形式，导致知识难以被理解，无法被模仿。2，缺乏“主观 Tokens”（比如需求、情绪和利弊）。所以大模型的推理过程是有缺陷的。

图灵奖获得者，深度学习三巨头之一的 Yushua Bengio 教授，认为实现通用人工智能最重要一步就是：把神经网络和因果推理相结合起来。事实上，我们的方案已经实现了这一点：记忆网络就是全连接的神经网络，从输入 Tokens 到被激活的“世界模型”，就是对客观世界组织方式的因果推理；从输入 Tokens 到被激活的“主观 Tokens 组合”（代表需求、情绪和奖罚的 Tokens），就是客观世界和机器自身需求之间的因果推理。

所以，我们实现了“把神经网络和因果推理相结合起来”。事实上，目前的大模型已经实现了客观推理能力和部分主观推理能力，但它们的推理过程，人类难以理解，无法模仿，所以难以被利用。

5.3 我们实现了层次化的“通用决策能力”

机器只强化学习一种任务：“如何满足自身需求？”，也只处理一种任务“如何满足自身需求”？所以我们的机器，决策是“面对自身需求”，而目前其他 AI 方案，决策是面对形形色色的“任务本身”。

信息输入，产生各种联想，有好有坏。降低那些带来“惩罚”的 Tokens 发生概率，提升那些带来“奖励”的 Tokens 发生概率，这就是“通用决策”！这和人类决策是相似，所以通用！

有了奖励函数的前置化和步骤化，机器就有了“决策能力”。有了“趋利避害”这个通用目标，机器就可以实现“通用决策”能力。

（5.3.1） 目前的“机器学习”不是真正的“机器学习”

面对一个新任务，人是根据自己的经验，预测不同决策的“好坏”，最多选几个方案去尝试。

面对一个新任务，目前机器靠强化学习，就是“不断试”，要么是（1）尝试一百万次，看结果（Google 各种打游戏的 AI）；要么是（2）请人类告诉我好坏（GPT-4，大模型，RLHF）[23]，然后才能获得处理这个问题的决策知识。

所以目前的机器学习，走的是“先尝试”+“再淘汰”的路子。所以他们应该叫“机器进化”，不能叫“机器学习”。所以我们提出了 AGI 需要真正的“机器学习”。

什么才是真正的“机器学习”？我们认为，真正的机器学习，应该像人类一样，面对一个新任务，可以根据自己过去的经验，来预测不同决策路径下的“好坏”，最多选有限几个方案去尝试，就可以获得处理新任务的决策知识。更进一步，我们认为真正的学习，也应该和孩子学习方式类似，通过语言来直接获得人类已经积累的经验。在面对新任务时，一次尝试都不需要，直接一次成功！比如在实验室里，老师教孩子们做实验时，是通过语言传授，直接把人类已有的决策经验传递给孩子们。孩子们可以在获得老师传递过来的知识后，可以在不同的环境下，利用语言获得的经验，和环境互动，就可以直接完成实验。尽管孩子们可能是第一次做这些实验！

真实任务千差万别，真实场景千差万别，人类无法把每一类任务都放到大量场景中去“强化学习”！所以，必须转变思路！思路就是：把所有任务都转换为单一任务：“如何满足自己的需求”？机器的所有训练过程，都是训练这一个任务。所以，面对这个任务，机器已有大量的“state” and “policy”知识，所以可以预测“不同决策”下潜在的“利弊”估计。

而人类赋予给机器的任务，就是机器解决“如何满足自己的需求”任务的背景信息。如果“获得人类认可”是机器的需求之一，那么机器在追求“满足自己的需求”的过程中，就会把“完成人类的任务”纳入整体的利弊统计中。这和人类是类似的，面对老板给予的任务，你会全面权衡利弊，来做出不同的决策。比如

你做出的一种决策可能是：主动寻找更多的信息，来分析任务带来的利弊影响，再做决策。而主动寻找更多的信息，这就是人类自己给自己安排的新任务。如果机器也这样决策，那么就等同于机器给自己安排任务，也就是说，机器给自己编程了。事实上，我们的机器就是采用这样的决策流程：权衡利弊来定策略，并且有可能主动寻找信息来帮助自己实现利益最大化。

（5.3.2） 如何实现是真正的“机器学习”？

十年前，我们认为要想创建真正的“知识”，应该从信息统计的角度入手。不同于“深度学习”，我们认为机器应该按照人类学习模式，采用小样本，知识积累的方式来学习。所以，一开始也是试图走“符号表达”→“因果逻辑”→“知识网络”。

尝试几年后，发现这条路的第一步就走不通。因为“符号表达”→“狗”怎么表达？需要把“狗”的所有特征挑选出来。但“狗”可以是一个动物，也可以是一个人！可以是“一种被歌颂的性格”，还可能是“一种被鄙视的性格”，在不同的语境下，符号“狗”的含义差异极大。所以“狗”的本质是“狗”和其他所有事物关系的总和。所以“狗”，必须放到整个知识网络中，通过它和其他所有知识的关系来定义。所以，“符号主义”走不通！因为“狗”不能从其他知识中分割出来！必须建立类似于深度学习的“全连接知识网络”，这是我们的第一个结论。

因为“狗”，必须放到整个知识网络中，通过它和其他所有知识的关系来定义。所以必须要有足够的知识，才能把“狗”这个事说明白。所以，“知识数量必须要足够“，这样才能通过足够的背景知识来理解什么是狗。这是我们的第二个结论。

我们回头一看，这不就是大模型干的事吗？“深度学习”就是干全连接网络这事，大模型就是干“使用大量知识，来建立全连接知识网络”这事。

那么，为什么我们没有看到满大街走动的机器人？因为只有知识网络还不行！机器还必须能够“和环境互动决策”！有研究表明：人类每天都做 3 万多次决策。目前业界已知的，除了专家系统外，能让机器自己来实现决策，只有强化学习算法了。

所以，要想走向通用人工智能，一条可能道路就是：大模型 + 强化学习算法。事实上，GPT-4 已经实现了“全部知识 + 全连接网络 + RLHF”，RLHF 就是强化学习。Google 在 2022 年发布了 GaTo 模型，已经走了“全部知识 + 全连接网络 + 强化学习”道路。

那么，为什么我们没有看到 Google 推出满大街走动的机器人？

这条路的核心障碍是，强化学习算法，需要的两个前提条件[24]：

（1），机器需要知道不同决策路径下，它能获得的奖励信息。因为实际过程中，奖励信息存在稀缺和滞后的问题，所以目前解决这个问题靠大量的试错训练。（2），机器需要遍历搜索所有可能的决策。

这两个条件，在游戏里能完美满足。游戏可以不断试，决策的搜索空间有边界（还可以各种修剪降低搜索空间）。但现实生活中，很多问题无法不断试错（比如照顾孩子，没有人愿意让你不断试！），也没有明确的边界，所以这个问题解决不了！这就是 Google 不断推出可以打各种非常复杂策略游戏的 AI，却一直无法推出最基础的“家庭保姆型机器人”的原因！事实上，在日常生活中，绝大多数的决策复杂度，远没有游戏中的决策复杂！但因为现实生活中，很多事不能海量试错！而且现实生活中，相关的信息并没有明确边界。所以上面两大困难，导致 Open-AI 或者 Google，通过 “大模型 + 强化学习”，基本只能用来搞那些可以海量试错的东西。因此 AIGC，距离 AGI，还有很长的一段路！

我们的决策方案，本质也是强化学习，但只强化学习如何趋利避害。而且我们利用了在链式联想激活过程，自动限定了搜索范围！只搜索被激活的信息！而且我们利用了“Tokens”→“奖罚符号”的逻辑链，自动预测奖罚信息，而不是只有事后反馈才能获得奖罚信息。所以我们完美的解决了 Google 的决策型人工智能只能打游戏的问题！

这是因为我们同步实现了“客观常识”+“主观常识”。而现有的技术路线，采用的技术路线是先实现“客观常识”，然后通过“RLHF”来建立“主观常识”。所以目前的技术路线，“主观常识”是通过事后反馈来获得的，所以它只能适用于可以大量试错的领域。

（5.3.3）“通用决策”的实现过程。

机器在任意环境中，输入信息都包括所有传感器信息。所以在任何时刻，机器所处的环境信息都是输入信息的一部分。

机器和环境互动决策，包括两个方面：

1，最优决策的选择。

2，决策过程的执行。

这两个步骤，不是分开的！是交织在一起，并行处理的！

“通用决策”需要解决的第 1 问题是：奖励函数是什么？在 GPT-4 里面，在 Alpha go 里面，奖励来自于最终的外部反馈。而在我们的 AGI 中，奖励来自于外界信息所激活的“奖励”和“惩罚”符号，大小就是它们的激活值。

第 1 步：目的是啥？

当信息输入（外界 + 机器自身监控信息）输入后，有一些奖励符号和惩罚符号被激活。

每一条从输入 → 奖励符号和惩罚符号的激活值传递路径，就是一条潜在的，产生奖励或者惩罚的逻辑链路。

如果这条逻辑链路上，每一个底层特征都真实地实现了，那么这条逻辑链路所传播的奖励或者惩罚也就实现了。

所以机器对任何输入信息的响应，都一样：增加奖励逻辑链发生的概率，降低惩罚逻辑链发生概率，来达到趋利避害的目的。

第 2 步：有了目的，怎么规划？

1，怎么增加奖励链路，降低惩罚链路的发生概率？

就是增加，或者降低，链路上的高激活值 Tokens 组合的实现概率。链路上的高激活值 Tokens 组合，就是这条链路的高权重的 Tokens 组合。当它们为真，则沿这条链路传播的激活值为真，所以最终被激活的奖励，或者惩罚，也就为真的。

2，具体怎么操作？

从输入信息→奖罚符号的激活值传递路径上，选用激活值最高的 N 个 Tokens，它们就是导致奖励，或者带来惩罚为真的顶层实现路径。机器的目标就是：1，让奖励路径上的 Tokens 实现（就是模仿过去的经验，让它们出现在输入信息中）。2，让惩罚路径上的 Tokens 不能实现（就是模仿过去的经验，避免它们出现在输入信息中）。

所以，从输入→奖罚的逻辑通路上，选取激活值最高的 N 个 Tokens，包含它们的激活值传播路径，就是顶层实现路径。为什么机器只选激活值最高的 N 个 Tokens？因为这些 Tokens，要么是因为它们是一类事物的代表性 Tokens，所以记忆值高，从而获得了更高的激活值；要么就是和输入信息关系密切的 Tokens。由于数量少，相当于属性限定少，所以和它们关系最密切的概念通常是“抽象概念”。

由于语言符号使用很频繁，所以语言 Tokens 常常获得高激活值，成为构成“抽象概念”Tokens 组合的激活值最高的核心 Tokens，使得语言符号成为概念本身的代表。比如“吃饭”、“逃避”等抽象概念。需要指出的是，“抽象概念”并非是语言符号的专利，动物同样可以有“顶层决策”。

所以机器建立决策的过程，是优先“抽象概念”，然后逐步增加更多的 Tokens，形成更加具体的概念组合。这就是自顶而下，逐步展开的决策和执行过程。我们把这个过程称为“分段模仿”。

关于分段模仿的方法具体示例：

假定把输入 Tokens 的集合作为 A，把响应 Tokens 的集合作为 B;机器通过 A 和 B 链式联想激活过程，寻找那些高激活值 Tokens，这些 Tokens 是和 A、B 都存在连接关系的 Tokens，因为它们从 A 和 B 都获得激活值，所以成为高激活值 Tokens。它们就是连接 A 和 B 的中间桥梁 Tokens。这个过程迭代进行，就能实现自顶而下，逐层决策。

在计算机中怎么实现？采用的方法是：（1）外界输入 Tokens 进行链式联想激活过程→确定奖罚符号的激活值（超过预设值的 Tokens 作为目标），建立一级目标。（2）从激活值最高的奖罚符号开始，找到从输入到每一个一级目标的激活值传递路径上，激活值最高的 N 个 Tokens，它们就是实现对应奖罚的逻辑链路。链路上的 Tokens 就是二级目标。（3）机器以每个二级目标作为新目标，把它们作为一种新的输入 Tokens，给与它们初始激活值，再次发起链式联想激活过程。所以，那些最高激活值的 Tokens，就是和外界输入 Tokens，以及和二级目标 Tokens 都相关的 Tokens 组合。这是因为我们采用了激活值累计和激活值消退，只有和最近输入 Tokens 相关的 Tokens 才能维持激活状态。所以这些 Tokens 就是三级目标。（4）这个过程迭代进行，机器就能把每一个一级目标都分解成实现它们的层次化逻辑链路。（5）决策过程的每一次展开，会有不同的奖励值或者惩罚值被选择，进入累计。机器按照趋利避害的原则，选择带来奖励值的子路径，避免带来惩罚的子路径，从而增加累积的奖励值。当机器发现总的奖罚值收敛了，也就是无法进一步改善了，也就是利益最大化了。机器就停止进一步展开，进入执行过程。这就是 Yann Lecun 教程提出的层次化“通用决策能力”，也是 Bengio 教授提出的和神经网络兼容的逻辑推理能力。

为什么每一次展开都只选 N 个最高激活值的 Tokens？这是因为过去经验和目前实际不可能完全匹配，所以通过只选用最高激活值 Tokens，意味着由它们组成的“模型”要么是抽象的（适用范围广），要么是和输入 Tokens 密切相关的（匹配度好）。只选 N 个最高激活值的 Tokens 的目的，是为了实现经验泛化。所以，在我们的方案中，经验泛化是自动实现的。

比如，机器有使用钉锤砸钉子的经验，在需要砸钉子，并且没有钉锤的情况下，并且输入 Tokens 中存在石头的情况下，为了实现一级目标（奖励符号或者惩罚符号，完成任务，获得奖励，或者避免被惩罚），在被激活的逻辑链路上，可能包含了代表钉锤的 Tokens 组合。那么，这些 Tokens 组合就成为二级目标。

机器根据记忆库中的链式联想激活过程，可能发现了 M 条实现钉锤目标的激活值传递路径，可能是从“记忆中的工具箱出发”，还可能是从“向队友借用相关经验”出发，这些激活值传递路径都是提高“钉锤”Tokens 实现概率的路径，也就是通向奖励的二级路径。

由于石头相关 Tokens 则出现在输入中，钉锤和石头的共有 Tokens（比如重量数据、尺寸大小、硬度感觉等）就可能获得更高的累计激活值，从而被作为前 N

个高激活值，被挑选出来。它们就变成了桥梁 Tokens，使得从石头相关的 Tokens 出发，也成为通向奖励的二级路径。这就是经验泛化过程，通过石头和钉锤共有的 Tokens，使得石头的 Tokens 可以向奖励符号传递激活值。之所以能够实现，是因为“石头”和“钉锤”拥有部分共有属性（共有 Tokens，而这部分 Tokens 在记忆中，能够重复出现在各种“用钉锤砸钉子”的场景中），它们就是经验泛化的桥梁。可以看到，在我们的方案中，经验泛化过程是自动完成的。

那么，机器选用哪条通向奖励的二级路径呢？这时，机器需要根据新链式联想激活过程更新了的激活 Tokens 空间，按照行趋利避害原则，重新来选择自己的决策路径。有些路径，既可能带来奖励，有可能带来惩罚，这时导致机器在统计奖罚值时难以收敛。如果机器发现奖罚值统计没有收敛，机器的决策就是进一步识别信息，来收敛每一条奖罚值传递路径。

比如“记忆中的工具箱出发”就需要确认一下“工具箱”目前出现在输入中（实现）的概率是多少？这个概率就可以进一步收敛这条路径的奖罚值。这时，确认一下“工具箱”目前出现在输入中概率，就成为机器自我创建的一个新目标。为了完成这个“新目标”，机器需要模仿过去的经验去执行。假如在过去的记忆中，它的工具箱都挂在腰间，那么它模仿过去的经验去确定这个工具箱相关的 Tokens 出现在输入中，最可能模仿的过程就是用“手”拍一下腰间，去重现过去手碰到“工具箱”的各种传感器数据组合。因为这种决策下花费的电量最小，使用的时间最少，能够实现机器自身的利益最大化，所以这就是优选决策路径。

再比如，“向队友借用”这条路径，有需要提升这条路径上 Tokens 的实现概率，才可能给奖励符号传递更大的激活值（获得更大的奖励）。所以，机器最可能模仿的经验就是扭头看，或者询问。

所以，在我们的方案中，机器的决策是非常复杂的，在一个决策路径中，可能嵌套 W 个决策和执行过程，但任何时刻，机器的唯一目标都是“趋利避害”。所有决策都是围绕这个目标衍生出来的。所以，机器的决策是非常灵活的，它时刻根据环境状态变化而变化，并没有预置流程。而唯一的预置流程只是：“趋利避害”。

上述过程迭代进行，每一次都有新的奖罚符号被激活。机器通过统计这些奖罚符号的激活值，直到发现奖罚符号的激活值收敛为止。这时机器就建立了最优响应路径。

机器的决策有可能是对输入信息的响应，也有可能是寻找更多的信息来继续做决策。无论是哪一种，机器都是通过模仿过去的经验来提升或者降低特定 Tokens 的实现概率。任何时候，有新的信息输入后，新信息会通过链式联想激活过程，更新记忆库中的激活值分布。这时机器需要根据新的状态，重新统计奖罚信息，重新寻找最优决策。只有有新的信息，这个过程时时刻刻都在进行。

第 3 步：有了规划，怎么执行？

执行，就是通过模仿过去的经验，来提高，或者降低特定的 Tokens 发生概率。

1，选用少量最高激活值底层特征→抽象决策路径。

2，增加更多高激活值底层特征→抽象决策路径具体化。

3，上面步骤 1、步骤 2 迭代进行，直到把决策分解到可以执行的驱动命令为止。驱动命令: 给喇叭送波形，给电机发驱动命令，给显示屏发送显示数据，给表情展示系统发送设置参数等。

4，随时都可能碰到新输入信息，新输入信息会改变记忆库中的激活值，改变奖罚情况，所以机器在实施最优响应路径过程中，有可能随时改变原有计划！

第 4 步：决策和执行过程中的分段模仿。

机器通过链式联想激活过程，可以找到和目前输入相关的经验。这些经验中，那些少量高激活值 Tokens 构成的概率，由于抽象性高，所以它们通常是具有代表性的抽象经验。这些经验包含了和输入 Tokens 相关的“前因”和“后果”，它们就是经验泛化的对象。

经验泛化，本质就是利用已发生过程的因果，来实现未发生过程的因果。而在我们的方案中，是通过两个过程中“共有 Tokens”的激活值传递过程，来自动完成的。由于两个过程中，Tokens 并不一致，这就对应了经验泛化过程中的不匹配问题。但这个问题，在我们的方案中，两个过程的经验，是通过它们共有的 Tokens 来实现激活值传递过程，来自动完成泛化的。

需要特别指出，机器的概念是由各种 Tokens 通过庞大的立体网络形成的。同样的 Tokens 可能分布在不同的记忆片段里。这些 Tokens 既有可能来自于自身的经历，也有可能来自于语言符号的输入。

因此语言符号本身被激活后，会激活那些语言符号所代表的相关 Tokens。而语言符号本身存在次序，以及语言序列中通常还包含有表达 Tokens 组合次序的符号，所以语言符号序列的背后，是 Tokens 时间、空间信息流。它们的时间、空间组合次序，就是“因果关系”。并且，这些“因果关系”通过“语言符号”的链式联想激活过程，能够形成紧密的激活值传递关系。而这种紧密的激活值传递关系，本身就是一种“经验”。所以，在我们的方案中，经验不仅仅是来自于机器自身的经历，还来自于通过语言符号获得的“其他人的经验”。所以，我们的机器，既可以通过语言符号来学习“经验”，也可以通过语言符号构成的 Tokens 信息流，来模仿“经验”。

6，我们方案和目前大模型道路的对比。

我们的方案，解决了如下问题：

（1）如何“建立常识”的问题。

深度学习破坏了原有 Tokens 的时间、空间关系！而我们的方案中，采用“链式联想激活过程 + 记忆和遗忘机制”同样实现了注意力机制。但我们没有采用深度学习，所以我们的方案，其创建的知识保留了的 Tokens 原始的时间、空间关系。而原始的 Tokens 组合方式，正是人类“概念”的基础。所以，在我们的方案中，其创建的“知识”是人类可以理解，可以模仿的知识。

在我们的方案中，“知识”的本质就是 Tokens 在时间和空间中的排列关系，以及不同 Tokens 排列对智能体潜在利弊的预测。而 Tokens 在时间和空间中的排列关系本质就是“因果”，这些 Tokens 在时间和空间中的排列关系并非简单的时间、空间临近关系，而是智能体从中总结出来的，能重复出现的关系，它们实际跨越的时间和空间跨度有可能很大，但通过链式联想激活过程，这些时间和空间跨度大的 Tokens，形成了紧密的激活值传递关系，这就是知识。如果知识中包含了代表“需求”、“情感”、“利弊”相关的 Tokens，这就能预测潜在的利弊，所以 Tokens 的排列就代表了“知识”。而那些常见的排列就是“常识”。

（2）“机器是否可以有意识”的问题。

我们解决了如何给机器赋予“自我需求”。所以，机器能自主决策，自我进化，可以有自己的情感，可以追求“自我需求”，所以我们的机器是有“意识”的。

（3）“通用决策”问题。

机器面对任何任务，都按照“趋利避害”来决策。人类给与的任务，是机器追求“自我需求”的副产品。

这个和你完成老板交代的任务是一回事。你也是在追求“自我需求”的过程中，完成老板交代的任务。如果两者有冲突，你也会按照趋利避害来做出各种不同的变通决策，试探老板的真实意图，考虑老板的底线，所以你的决策会很灵活！

（4）“语言理解”问题。

因为我们没有破坏 Tokens 原来的时间、空间关系。语言序列代表的 Tokens 时间、空间序列是可以被理解，可以被模仿的。所以机器可以通过语言，像人类一样直接学习各种技能。读一遍烤箱手册，就可以开始烤面包[5][6][7]。

我们认为，我们的道路，是一条通向 AGI 的可行道路。

优势 1：能处理那些无法大量试错的任务。

比如自动驾驶，家庭保姆，照顾老人，陪伴孩子，从事“工农兵学商”。

因为我们是“类人”AI，能够通用决策，能用语言学习技能！而目前大模型无法搞定这些事！

优势 2：能解决“幻觉”问题。

大模型只有局部统计获得的“常用语”，没有事实记忆。

我们的方案，首先是存储记忆，然后从记忆中提取常见信息。所以我们是自带“事实数据库”，而且和知识融为一体。

优势 3：能通过语言直接学习技能，并模仿。

因为我们没有破坏 Tokens 组合的时间、空间关系，所以语言代表的 Tokens 时空关系可以被理解，被模仿！这一点，无论现在，还是未来，大模型都实现不了！比如“机器人”第一天到面包店上班，它会找老板要“烤箱”操作手册。读一遍，直接开始“烤面包”，不需要单独的训练！

优势 4：更安全！

（1）目前人工智能是单一目标，从决策来讲，它就是“为达目标，不择手段”的“一根筋思维型”人工智能。这样的人工智能，它不会去考虑目标之外的任何东西，决策还是黑盒的。想想如果“闷罐子”+“一根筋”类型的人控制了你的生活，这有多危险！如果让这样的人工智能全面掌控人类的生活，它完全可能因为理解错误，出于好意地给人类带来无法估量的灾难。

（2）而在我们的方案中，机器的“需求类型”是可以预置的，价值观是可以被训练的，可以对齐人类的价值观，任何时候机器都会综合考虑各种目标，不会出现“偏激”行为。而且，在我们的方案中，决策是可见的、可修改的，是“白盒”的。

7，我们方案的底层逻辑。

7.1 链式联想激活过程就是注意力机制。

首先，我们认为知识的本质是信息。而人类产生的知识，是信息的极小一部分。这是因为，我们人类对信息的分辨率是有限的。一颗小草上 A 原子和 B 原子排列的相对时空关系，也是一种信息，但我们不会去识别它。

所以人类在进化的过程中，产生了 Tokens 识别能力。Tokens 就是人类常用的最小信息单元，比如一根直线。Tokens 本身就是“世界模型”，它是人类用于搭建宏伟的知识殿堂的最小“世界模型”。人类在进化过程中，形成了采用 Tokens

这样的“模型”来识别周围信息的“模式识别”能力，极大的提升了信息识别的能效比。这是进化带给我们的礼物。

如果我们把从“宇宙大爆炸”到“现在”，所有事物的“Tokens”，按照空间、时间次序排列起来。我们就获得了一个信息张量。它就是人类拥有的全部知识。

面对这样的知识宝库，如果我们宇宙之外的智能体想了解它，它们会对这些 Tokens 做统计。

第一个问题：“我们有多少种独立的 Token”？在我们的方案中，相似性关系回答了这个问题。第二个问题：“每一种 Token 的数量分布”？在我们的方案中，重复性关系回答了这个问题。第三个问题：“Tokens 之间是怎么排列的”？在我们的方案中，临近性关系回答了这个问题。我们可以看到，在我们的方案中，通过链式联想激活过程，记忆和遗忘机制，就是对信息做统计学意义上的描述！

在大模型的注意力机制中，通过 Tokens 之间两两相关性，来推测 Tokens 组合相关性。然后再次通过两两相关性，来推测更大 Tokens 组合相关性。这个过程经过多次迭代，就能获得不同 Tokens 组合彼此之间的相关性。而预训练过程，就是通过试错法（深度学习），来寻找正确的“最优坐标基底”。在注意力机制的帮助下，所获得的“最优坐标基底”只是针对“常见信息”而言。这个过程本质就是贝叶斯推理过程：通过部分已知概率，来推测某种特定 Tokens 的条件概率。

在我们的方案中，Tokens 之间的相关性，是通过归纳法来获得的。链式联想激活过程，是利用预训练获得的相关性（部分已知概率），来获得某种特定 Tokens 可能出现的条件概率。而链式联想激活过程，就是寻找相关性（注意力机制的推理过程）；而记忆和遗忘机制，就是一种归纳法。

7.2 注意力机制的核心就是创建“常识”。

知识就是 Tokens 的排列方式，常识就是常见 Tokens 的排列方式[16]。目前大模型的核心问题就是它把人类的知识（Tokens 的排列方式），转化为了它自己的一套知识体系（因为深度学习破坏了原有 Tokens 的时空关系，导致大模型的知识，人类难以理解，无法模仿），它用它自己的知识体系来解决问题，然后再翻译给人类。所以，深度学习破坏了原有 Tokens 的时空关系，是指它破坏了 Tokens 原有的、人类可以理解的组织形式，而转换成了机器可以理解的组织形式。从机器的角度看，它保留了 Tokens 的组织方式，因为它正确的找出了“常见信息”。但从人类的角度看，它产生的知识，和人类创建的知识，无法直接互联互通，无法直接相互借用。

因为两套体系的底层语言彼此无法沟通，所以人类难以给机器赋予“先天知识”（比如先天需求、先天奖罚函数和先天情绪函数），所以只能采用后天补救的方式，采用 RLHF，或者采用外挂知识库，来解决部分问题，而且只能通过“yes” or “No”来沟通，这样的机器人，只能是一个“照本宣科”的“书呆子”，无法真正地灵活解决问题。

所以，在我们的方案中，最核心的是要不破坏 Tokens 原有时间、空间组织形式下，建立“常识”，并且需要包含机器的“主观常识”。

为了不破坏 Tokens 原有时间、空间组织形式下，建立“常识”，我们采用了信息 Tokens 化，并保留时间、空间信息存储，并采用了链式联想激活过程，并采用了记忆和遗忘机制来实现 Tokens 之间的链式激活值传递关系的归纳。同时，我们模仿“常识”的组织形式，预置了代表先天的需求、先天奖罚函数和先天情绪函数的 Tokens 组合。然后让机器按照趋利避害的原则，自主决策，自我进化，

围绕先天知识不断扩展记忆库，形成整个知识网络，从而创建“客观常识”和“主观常识”。

7.3 我们只完成一件事：“创建常识”。

为了“创建常识”，要先解决了（1）“给机器赋予自我需求”。

为了解决（1），就要先解决（2）“如何创建能理解的知识”问题。

为了解决（2），要解决“不使用深度学习，如何创建全连接知识网络”的问题。然后就可以实现主观 Tokens 和客观 Tokens 通过注意力机制，建立连接关系。这就是常识。

主观 Tokens 和客观 Tokens 建立关系，就是励函数的“前置化”+“步骤化”，就能通过“趋利避害”来实现“通用决策能力”。在“自我需求”的驱动下，机器就能实现“自我进化”。

7.4 我们建立一个婴儿 AI。

“建立一个婴儿机器，然后终身学习，自我成长”。这个想法已经很多年了，但我们是第一个提出详细解决步骤的团队。

8，一个简单的示例

下面，我们通过一个例子，来说明机器如何决策和响应。

背景：老王去外地度假，带了一个助理机器人，住进了酒店房间...

老王：“喂...”。

机器人：记忆库中有很多 Tokens 处于激活状态，但这些被激活的 Tokens 里面，没有激活值超过 A1（A1 是一个预设阈值）的奖励符号，也没有激活值超过 P1（P1 是一个预设阈值）。

它处于持续接收传感器传来的外部信息和内部信息，并采用低分辨率优先提取这些信息中的 Tokens，存入记忆库。按照同样的流程，给这些 Tokens 赋予初始激活值，由于没有高激活值的奖励符号 / 惩罚符号，所以按照预定程序，给这些 Tokens 赋予的激活值比较低，所以在随后的链式联想激活过程中，激活值传播范围很小，链式激活过程很快完成。

机器开始更新记忆值。由于被激活的 Tokens 获得的激活值低（因为初始激活值低，激活值传播范围小），所以它们增加的记忆值很小，很多信息短时间就会被忘记。同时，由于记忆库中的奖励符号、惩罚符号获得的激活值都比较低，也就是潜在的奖励，和潜在的惩罚都比较小。所以机器形成的最佳决策路径就是继续接受信息。这是因为付出电量本身是一种惩罚，如果没有获得奖励，那么最优决策就是不浪费电量。

每一次链式联想激活过程完成后，机器都需要查看有没有激活值超过预设阈值的奖励或者惩罚符号。这种情况下，机器形成的最优响应就是：采用低分辨率提取这些信息中的 Tokens，存入记忆库。按照同样的流程，上述过程循环进行。

突然，音频处理系统传入了一连串音频 Tokens（依然采用低分辨率提取的），这些 Tokens，按照同样的流程，被赋予比较低的初始激活值，并进行链式联想激活过程。这次输入的 Tokens 中，有些 Tokens 在链式传播过程中，因为相似性，激活了记忆库中很多相似的 Tokens，这些 Tokens 和很多奖励、惩罚符号之间存在紧密的激活值传递关系，所以这一次进行的激活值链式传播过程，有很多奖励和惩罚符号被激活了。（这些 Tokens 通常就是主人的声纹特征，比如特有的音色）。

由于这一次有很多奖励、惩罚符号获得了超过预设的激活值。假设有 N 个奖励符号和 M 个惩罚符号的激活值超过预设值。机器以 N 个奖励符号为目标，也以 M 个惩罚符号为目标，这样机器自主同时建立了 N+M 个目标。所以，在我们的方案里，目标是机器自主产生的，是同时产生多目标的，而不是人为预设一个总的奖励函数。

在我们的方案中，机器的一切响应，都是以趋利避害为原则。所以机器创建 N+M 个目标后，机器规划自己的响应路径原则是：提高奖励符号的激活值发生概率，降低惩罚符号的激活值发生概率。所以机器的决策，就是围绕实现奖励，避免惩罚展开的。

机器首先处理激活值最高的那些奖/罚 Tokens，可能是一个或者多个惩罚 Tokens；在记忆库中，向这个惩罚 Tokens 传递激活值的传播通路可能是：主人的声纹底层特征输入，通过相似性激活率记忆库中很多主人的声纹特征；这些激活值在记忆库中进一步链式传播激活值。

在这些记忆中，有一个惩罚 Tokens 的激活值很高。而能获得高激活值的 Tokens，无非就是几种 cases：（1）这个惩罚 Tokens 的记忆值很高。一种可能的原因是存储它时，它的激活值很高，而记忆值增量和激活值正相关。另外一种原因是它常常被激活，通过重复获得了高记忆值。（2）多个输入 Tokens，通过不同的路径向这个惩罚 Tokens 传递了激活值。比如主人的“语气 Tokens”，“用词 Tokens”，“主人的状态 Tokens”，“主人的表情 Tokens”、“目前环境相关 Tokens”等，如果这些 Tokens 都和类似的惩罚符号在记忆中存在紧密激活关系，那么它们一起完成激活值链式传播过程后，和它们都相关的 Tokens 就可能获得高激活值。（3）这个惩罚 Tokens 和特定的输入 Tokens 之间存在紧密激活值传递关系。也就是说，它们在记忆中，总是伴随出现。所以它们之间形成了“临近关系”和“高记忆值关系”，并且传播路径很短，激活值传递系数高。所以注意力机制，既可能通过综合推理（比如多个 Tokens 向特定奖罚符号传递激活值，综合经验），又可能采用特例推理（比如特定的激活值紧密传递路径，特定经验），形成了对信息的注意机制推理。

奖罚 Tokens 的激活值高，还可能来自于之前的 Tokens 输入所建立的激活值分布。尽管高激活值 Tokens 的激活值会随时间而消退，但如果激活值足够高，它将在更长时间内影响机器的决策。这和人类很相似。

在这个例子中，激活值传播路径构成的传播网络包含的 Tokens 非常多，难以表述。但通常是语言符号的激活值最高（因为它们最常用，记忆值最高），如果这些语言符号按照它们的时空次序组合起来，大意可能是“不要躺着（前因），被主人骂了，很难过（后果）”。

于是机器立即开始搜索最优响应路径，用于避免这个惩罚符号发生的概率。机器做决策的原则是提高奖励符号发生概率，降低惩罚符号发生概率；而具体采用的方法是：针对传播激活值给惩罚符号的路径上的概念，降低它们发生的概率；针对传播激活值给奖励符号的路径上的概念，增加它们发生的概率；对这些概念又具体如何增加、降低概率呢？每个概念是记忆库中局部紧密网络，机器需要降低这个局部紧密网络中高激活值 Tokens 发生的概率，从而降低这个奖罚逻辑链路的发生概率。

比如在机器的记忆中，被主人“采用相似的 Tokens 训斥”时，记忆中存储了自己当时的内部传感器数据 ，也存储了当时外部传感器数据；其中一些 Tokens 因为后来没有再次重复，没有获得增强记忆而被遗忘了。但能和这个“惩罚符号”

共同重复出现的 Tokens 组合中，都有“躺着”相关的 Tokens、以及一些“代表特定时间 Tokens”、以及“代表特定场合 Tokens”，它们因为重复性，获得了更高的记忆值。并因为是能重复出现的组合，每一次彼此都推高对方的激活值，所以获得了远比重复性更高的记忆值。而且因为它们能重复，所以它们的组合，每一次都能获得更高的激活值，所以它们更加容易被激活，从而更加容易被记忆，所以这是一个正向循环过程。这就是经验总结过程。

如果有一次，在类似的环境下，主人却表扬了机器，这样的记忆后续也会被参与决策。所以，在类似的环境下，各种 Tokens 既可能向惩罚符号传递激活值，也可能向奖励符号传递激活值。所以，机器的决策是综合统计所有的奖罚值，既可能考虑如何获得奖励，又会考虑如何避免惩罚，所以机器在选择响应路径时，有些局部响应路径既是通向奖励的路径，也是通向惩罚的路径，所以机器需要对这些路径进行细分，来确定什么是通向奖励的路径，什么是通向惩罚的路径。而这个细分过程，就是给这个路径增加更多的 Tokens，从而形成多条细分路径（比如不同的场景下，或者不同的时间点，或者不同的前因等），这样机器就可以通过细分路径来确定自己的响应，这就是分段模仿的核心。

所以，我们的机器不需要通过修改过去的参数来容纳新的“Fine tuning”。它只是需要通过积累记忆来实现“Fine tuning”。它可以进行任何深度的“Fine tuning”，可以进行任何领域的“Fine tuning”，而且还可以进行无数领域叠加的“Fine tuning”，而不会发生“灾难性遗忘”。这是因为它并不会修改过去的知识参数，而只是简单的扩增网络。

在本例中，假设是白天，假设机器正躺着（节省点电，获得奖励），当主人的声纹激活惩罚符号后，机器需要避免被激活的惩罚符号发生概率，提高被激活的奖励符号发生概率。那么，这里至少有两种 Case，1，降低“躺着”概念的发生概率，避免惩罚（比如被训斥）；2，提高“躺着”概念的发生概率，获得奖励（比如省电）。这时机器就需要按照趋利避害原则，做出最优选择。这时，机器就要综合各种响应路径，对比统计奖罚值。

假设这时机器的电量充足，省电带来的奖励很小。在完成链式联想激活过程后，只有一个惩罚符号获得了高激活值。机器按照趋利避害原则，会选择避免惩罚，因为这样的统计下奖励值最高。所以机器在利益最大化驱使下，会把避免惩罚作为目标，开始建立响应。

假设这时机器的电量不足，省电带来的奖励很大（这里假设机器必须躺下充电）。在完成链式联想激活过程后，有一个惩罚符号获得了高激活值，还有一个奖励也符号获得了高激活值。机器按照趋利避害原则，会同时建立两个目标：实现奖励，避免惩罚。因为这样的统计下奖励值最高。所以机器在利益最大化驱使下，会把获得奖励 + 避免惩罚作为目标，开始建立响应。

假设机器的电量充足，那么现在，机器创建了第 2 级目标：降低“躺着”概念所获得的激活值。于是，在第 2 级目标的约束下，机器寻找向“躺着”概念传递激活值的传播路径，并创建第 3 级目标：降低这些传播路径上概念的激活值。于是，机器发现向“躺着”概念传播激活值的主要路径是一组自身状态传感器的输入。于是，机器创建了第 3 级目标：降低这些输入 Tokens 的概率。

机器会记录每一次训练的各种内外参数，采用记忆和遗忘机制来优化；通过奖罚反馈，鼓励机器模仿获得奖励的参数，避开获得惩罚的参数。通过这样的方式，参数组合+奖励+内外环境三者之间就建立了经验性的连接关系。这本质上是一个强化学习过程。当然，人类也可以模仿其形式，给机器置入先天的知识（驱

动相关），或者利用人类已经积累的经验，直接修改机器的知识，使其尽快收敛。

所以机器在不同的环境下，环境 Tokens 会自动激活最相关的记忆，通过模仿这些经验，向机器的运动系统传递相似的参数组合（包含参数类型和它们的时间次序，这些过程都是自动完成的）。这样机器就可以在各种环境下站立起来，降低“躺着”相关 Tokens 的发生概率。

假设这时机器的电量不足，机器实现奖励的经验会让它继续躺着，让充电相关的 Tokens 实现概率提升。而避免惩罚的经验，它会模仿过去的经验，给主人解释自己这么做的原因。然后机器创建了第 2 级目标：提升"充电"概念所获得的激活值。模仿过去避免“惩罚”的经验。所以机器可能创建第 3 级目标：“给主人解释自己行为的原因”，因为这样的“Tokens 组合”在记忆中，和“避免惩罚”这样的 Tokens 组合之间存在紧密的激活值传递关系，所以机器的目标就是提升和特定 Tokens 组合（给主人解释自己行为的原因）”的发生概率。所以下一级决策相关的 Tokens 组合就是：语言组织相关经验就会被激活。

这个过程迭代进行，每一次都有新的奖罚符号被激活。机器通过统计这些奖罚符号的激活值，直到发现奖罚符号的激活值收敛为止。这时机器就建立了最优响应路径。

然后，机器进入模仿执行过程。机器的决策路径，需要迭代分解到底层驱动参数为止，才能通过模仿经验中的参数配置，来发出驱动命令，从而模仿执行。

而在实际情况中，经验和现实总是只能部分匹配，所以经验和现实之间，也只能通过模仿它们共有的 Tokens 组合方式来实现泛化。

这些路径中，那些高激活值 Tokens 组成的路径，就是顶层模仿路径。如果模仿路径中不包含直接的底层驱动命令组合，那么增加更多的 Tokens（更低激活值 Tokens）进来，这时模仿路径就变成了更多 Tokens 形成的多段路径的不同组合形式。这就是分段模仿的含义。

也就是说，我们面对一个大的路径，没有合适的经验直接模仿，那么就细化，分解成多个小的响应路径段，针对每一个小的路径段，重新来寻找合适的经验来泛化经验。如果还是不能分解到直接的底层驱动命令组合，那么重复这个过程，通过增加更多的 Tokens，把这个响应路径分解成更多的小路径段，然后寻找合适的经验来泛化经验。如果还是不能分解到直接的底层驱动命令组合，那么重复这个过程，直到分解为直接的底层驱动命令组合为止。

上面过程不断迭代进行。时时刻刻都可能有新的 Tokens 输入。每当新的 Tokens 输入后，机器都需要再次进行链式联想激活过程。完成后，记忆库中的激活值分布发生了变化，所以机器需要重新进行决策过程。所以在这个过程中，机器的最优决策有可能是放下目前部分目标，开始追求最新产生的目标。

所以，我们的机器会产生自己的目标，并可以不断的改变自己的目标，所以它的决策是非常灵活的，是时刻和环境相匹配的。

所以在上面这个例子中，机器可能的执行结果是：立刻站起来，同时提高声音处理系统的分辨率，同时扭头去观察主人的姿态、动作和表情，但直到这个时刻，主人可能刚刚说完“喂...”字，后面的话还没有开始。

所以，我们的机器是类人智能，它对信息的理解，来自于它自身的经历，而不是来自于统计过程。也只有这样，我们的机器才可能有个性化服务。

一千个家庭主妇，有一千种不同的要求。通过知识统计获得的人工智能，无法实时更新知识的机器人，永远无法走进家庭，永远无法走进主妇们的心。它们的落地场景将非常有限，而我们的方案，才是真正的通用人工智能，它也许将改

变世界的面貌。

9，结束语

我们认为，人工智能的发展可以近似分为不同的阶段：（1）“特征探索”阶段。深度学习之前，主要集中在“人工探索”阶段。在深度学习之后，集中在“机器探索”阶段。（2）在实现了真正的注意力（Transformer）之后，因为机器的“知识坐标基底簇”和人类“知识坐标基底簇（概念）”初步对齐后，机器实现了“知识泛化”。面对人类的任务，机器可以通过“知识泛化”表现出一定的智能。

一维注意力机制，带来了语言大模型。二维注意力机制，带来了图像泛化。三维注意力机制，能实现 3D 创造能力。四维（三维+时间）注意力机制，能实现动态过程的泛化：会带来视频生成，也会带来限定场景下机器人服务。

但我们认为，只有增加“生命力：第五维，自我需求”，才可能给机器智能带来真正的“灵魂”。而大模型走的这条路，注定了它无法实现“第五维度”。而我们的方案能给机器赋予“生命”，所以它才可能成为真正的“通用人工智能”。

所以，我们认为人工智能需要发展到下一个阶段：“自主互动”阶段。“自主”意味着机器不再是沉默的“机器”，它能够自发地产生行为（这等同于给自己编程），机器会自我探索知识（比如主动和环境互动，获得知识）。“互动”意味着机器可以和环境实时互动，实时更新自己的知识，并能进行连续决策，在陌生环境下完成复杂的任务。

如何走向真正的通用人工智能，很多著名学者都提出了自己的看法，比如 Lecun 教授提出的“世界模型”，朱松纯教授也提出了实现通用人工智能的四个特征：（1）能够执行无限的任务；（2）能够自主生成新任务；（3）有价值系统驱动；（4）拥有反映真实世界的世界模型。显然，我们的方案，就是对 Lecun 教授、朱松纯教授思想的响应。

通用人工智能是人工智能的初心，也是人工智能的桂冠。我们提出了一套实现通用人工智能的技术方案，包含有 Step by Step 的实现步骤。在参考文献[25][26][27][28]中，我们通过专利的形式，详细揭示了实现这条道路的技术步骤。它也许将是一条引导人类走向通用人工智能的正确道路。

参考文献：

[1] A Generalist Agent, Scott Reed, Konrad Zona , Emilio Parisotto , Sergio Gomez Colmenarejo , Alexander Novikov, Gabriel Barth-Maron, Mai Giménez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar and Nando de Freitas, https://openreview.net/forum?id=1ikK0kHjvj

[2] ChatGPT 对科学研究和文献情报工作的影响,张智雄,钱力,谢靖等，chinaXiv: 202303.00093

[3] Ouyang L, Wu J, Jiang X, et al. Training language models to follow nstructions with human feedback[J]. arXiv:2203.02155, 2022

[4]Vinyals O, Ewalds T, Bartunov S, et al. Starcraft ii: A new challenge for reinforcement learning[J]. arXiv:1708.04782, 2017.

[5] A Survey of Learning-based Automated Program Repair, Antonio Mastropaolo,

Luca Pascarella, Emanuela Guglielmi, Matteo Ciniselli, Simone Scalabrino, Rocco Oliveto, Gabriele Bavota, arXiv:2302.00438
[6] A Survey of Learning-based Automated Program Repair,Quanjun Zhang, Chunrong Fang, Yuxiang Ma, Weisong Sun, Zhenyu Chen,arXiv:2301.03270
[7] Introducing ChatGPT, https://openai.com/blog/chatgpt/
[8] Prompting GPT-3 To Be Reliable,Chenglei Si, Zhe Gan, Zhengyuan Yang, Shuohang Wang, Jianfeng Wang, Jordan Boyd-Graber, Lijuan Wang, arXiv:2210.09150
[9] An experimental open-source attempt to make GPT-4 fully autonomous, https://GitHub.com/Significant-Gravitas/Auto-GPT
[10] https://jina.ai/news/auto-gpt-unmasked-hype-hard-truths-production-pitfalls/
[11] https://github.com/Torantulino/Auto-GPT?ref=jina-ai-gmbh.ghost.io
[12] Auto-GPT - The next evolution of data driven Chat AI, https: // auto -gpt.ai
[13] 文心一言：https://mp.weixin.qq.com/s/0-8X9FPouteKzNiK6DPaiA
[14] Learned in translation: Contextualized word vectors. In Advances in Neural Information Processing Systems
[15] Attention Is All You Need，Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin，arXiv:1706.03762
[16] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term memory-networks for machine reading. arXiv preprint arXiv:1601.06733, 2016.
[17] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio, A structured self-attentive sentence embedding. arXiv:1703.03130, 2017.
[18] Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model. In Empirical Methods in Natural Language Processing, 2016.
[19] Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304, 2017.
[20] Tamkin A, Brundage M, Clark J, et al. Understanding the capabilities, limitations, and societal impact of large language models[J]. arXiv:2102.02503, 2021.
[21] BERMo: What can BERT learn from ELMo?Sangamesh Kodge, Kaushik Roy,arXiv:2107.03508
[22] OpenAI. ChatGPT: Optimizing Language Models for Dialogue.https://openai.com/blog/chatgpt/, 2023
[23] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[J]. arXiv:2203.02155, 2022
[24] Principles of Solomonoff Induction and AIXI,Peter Sunehag, Marcus Hutter,arXiv:1111.6117
[25] 一种实现通用人工智能的方法， 陈永聪，曾婷，陈星月，中国专利

CN111553467B
[26] 一种模仿人类智能的机器智能实现方法陈永聪，曾婷，陈星月，中国专利 CN111563575B
[27] 一种实现类人通用人工智能机器的方法，陈永聪，曾婷，陈星月，中国专利 CN112215346B
[28] ESTABLISHMENT OF GENERAL-PURPOSE ARTIFICIAL INTELLIGENCE SYSTEM，陈永聪，张俊，曾婷，陈星月，美国专利号 11,715,291